0-315-03628-1

National Library of Canada
Canadian Theses Division
Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Division des thèses canadiennes

51523

PERMISSION TO MICROFILM — AUTORISATION DE MICROFILMER

- Please print or type — Écrire en lettres moulées ou dactylographier

Full Name of Author — Nom complet de l'auteur
JANIS MILDRED KYLE

Date of Birth — Date de naissance
JULY 17, 1946

Country of Birth — Lieu de naissance
CANADA

Permanent Address — Résidence fixe
10951 62 AVENUE
EDMONTON, ALBERTA
T6H 1N3

Title of Thesis — Titre de la thèse
AN EVALUATION OF NEW STATISTICS DESIGNED FOR ONE SAMPLE REPEATED MEASURES RESEARCH WITH MISSING DATA.

University — Université
UNIVERSITY OF ALBERTA

Degree for which thesis was presented — Grade pour lequel cette thèse fut présentée
M.ED

Year this degree conferred — Année d'obtention de ce grade
1981

Name of Supervisor — Nom du directeur de thèse
DR T.O. MAGUIRE

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

Date
APRIL 6, 1981

Signature
Janis M. Kyle

NL-91 (4/77)

National Library of Canada
Collections Development Branch

Canadian Theses on
Microfiche Service

Bibliothèque nationale du Canada
Direction du développement des collections

Service des thèses canadiennes
sur microfiche

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

Ottawa, Canada
K1A 0N4

NL-339 (Rev. 8/80)

THE UNIVERSITY OF ALBERTA

An Evaluation of New Statistics Designed for One Sample Repeated Measures Research with Missing Data

by

© Janis M. Kyle

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE
OF Master of Education

IN

EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA
Spring 1981

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR Janis M. Kyle

TITLE OF THESIS An Evaluation of New Statistics Designed for One Sample Repeated Measures Research with Missing Data

DEGREE FOR WHICH THESIS WAS PRESENTED Master of Education

YEAR THIS DEGREE GRANTED Spring 1981

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(SIGNED) J. Kyle

PERMANENT ADDRESS:

10951 62 AVENUE
EDMONTON, ALTA.
T6H 1N3

DATED FEBRUARY 2 1981

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled An Evaluation of New Statistics Designed for One Sample Repeated Measures Research with Missing Data submitted by Janis M. Kyle in partial fulfilment of the requirements for the degree of Master of Education.

T.O. Maguire

Supervisor

S. Hunka

George Sawada

Date. FEBRUARY 2, 1981

ABSTRACT

The purpose of this study was to examine the performance of seven statistics which have been developed to test the equallity of means in the one sample repeated measures research design with missing values on both measures. Six new statistics : ZLS, TLS1, and TLS2 developed by Lin and Stivers; ZE and SE developed by EKbohm; and TB1 developed by Bohj were compared to the paired t.

Each of these statistics was examined under the nine combinations of rho equal to .5, .7, .9, and delta equal to 0.0, .4, and .8. Nine different sample sizes, with varying numbers of observations missing on both variables were also examined. Populations of 3000 pairs of normally distributed variables were generated for each of the nine combinations of rho and delta. From each of these, 1000 random samples were selected for each sample size.

The six new statistics were evaluated and compared to the paired t in terms of a) the adequacy of fit to the respective t distribution, b) size: the proportion of probability values less than specified alpha levels when delta is equal to 0.0, and b) power: the proportion of probability values less than specified alpha levels when delta is greater than 0.0. Of the six new statistics, only three, TLS1, ZLS, and ZE showed promise as statistics to be used in preference to the paired t. Since the performance of

these statistics was basically acceptable with regard to fit and size, the major advantage of one statistic over another was due to larger gains in power over the paired t. Specifically, with rho=.5, TLS1 was the preferred statistic if the number of paired observations (n) was small (less than 16), and ZLS was the preferred statistic if n was large. With rho equal to .7, ZE, ZLS, and TLS1 were equally acceptable for small n, while ZLS was preferred for large n. Gains in power were minimal as rho reached .9, with only a slight preference for ZE when n was equal to eight.

Overall, it was concluded that only when the value of the population correlation was low and the true mean difference between the two population means was small, were any of the new statistics more powerful than the paired t. In such cases, the absolute gain in power was greatest for the small sample sizes.

Table of Contents

Chapter Page

I. INTRODUCTION 1

A. NOTATION 4

II. LITERATURE REVIEW 6

A. STATISTICS FOR DATA WITH MISSING OBSERVATIONS ON ONE VARIABLE 6

MEHTA AND GURLAND 6

MORRISON 8

NAIK 9

LIN AND STIVERS 10

EKBOHM 13

SUMMARY OF PROBLEMS 16

B. STATISTICS FOR DATA WITH MISSING OBSERVATIONS ON BOTH VARIABLES 17

LIN AND STIVERS 17

EKBOHM 20

BOHJ 26

III. METHODOLOGY 30

A. INTRODUCTION 30

B. PROCEDURES 31

CREATION OF POPULATIONS 31

SAMPLING AND CALCULATION OF STATISTICS 31

EVALUATION OF STATISTICS 32

IV. RESULTS 35

A. ADEQUACY OF DATA GENERATION PROCEDURES 35

POPULATION 35

SAMPLE 35

B. EVALUATION OF STATISTICS 37

GOODNESS OF FIT TO THE t DISTRIBUTION 37

SIZE 53

POWER 59

V. DISSCUSION, CONCLUSIONS, AND IMPLICATIONS 71

A. DISSCUSION 71

B. CONCLUSIONS 72

EVALUATION OF THE SIX NEW STATISTICS 72

COMPARISONS WITH THE PAIRED t 74

C. IMPLICATIONS 76

VI. BIBLIOGRAPHY 80

VII. APPENDIX I 83

LIST OF TABLES

Table	Description	Page
1	The most powerful statistic(s) for each condition of rho, delta, n, and j - from Lin and Stivers(1975)	12
2	The most powerful statistic(s) for conditions of positive rho and equal variances - from Ekbohm(1976)	14
3	The most powerful statistic(s) for conditions of positive rho and unequal variances when n=10 and j=5 - from Ekbohm(1976)	14
4	The preferred statistic(s) for varying conditions of rho and R(varX/varY) when n=5, j=5, k=10, and delta=1 and 3 - from Lin and Stivers(1974)	21
5	The preferred statistic(s) for varying conditions of rho and R(varX/varY) when n=10, j=5, k=10, and delta=1 and 3 - from Lin and Stivers(1974)	21
6	The preferred statistic(s) for varying conditions of rho and R(varX/varY) when n=15, j=5, k=10, and delta=1 and 3 - from Lin and Stivers(1974)	22
7	The preferred statistic(s) for varying conditions of rho and R(varX/varY) when n=20, j=5, k=10, and delta=1 and 3 - from Lin and Stivers(1974)	22
8	Characteristics of statistics	24
9	Goodness of fit of populations to the Normal Distribution	36
10	Kolmogorov-Smirnov p-values for the goodness of fit of the paired t, ZLS, TLS1, and ZE to their respective t distributions (delta=0.0)	38
11	Range of Degrees of freedom for TB1, TLS2, and SE	40
12	Kolmogorov-Smirnov p-values for the goodness of fit of TB1, TLS2, and SE : sample set 888	42

13 Kolmogorov-Smirnov p-values for the goodness of fit of TB1, TLS2, and SE : sample set 8816 43

14 Kolmogorov-Smirnov p-values for the goodness of fit of TB1, TLS2, and SE : sample set 81616 44

15 Kolmogorov-Smirnov p-values for the goodness of fit of TB1, TLS2, and SE : sample set 1688 45

16 Kolmogorov-Smirnov p-values for the goodness of fit of TB1, TLS2, and SE : sample set 16816 46

17 Kolmogorov-Smirnov p-values for the goodness of fit of TB1, TLS2, and SE : sample set 161616 47

18 Kolmogorov-Smirnov p-values for the goodness of fit of TB1, TLS2, and SE : sample set 2488 48

19 Kolmogorov-Smirnov p-values for the goodness of fit of TB1, TLS2, and SE : sample set 24816 49

20 Kolmogorov-Smirnov p-values for the goodness of fit of TB1, TLS2, and SE : sample set 241616 50

21 Size (Empirical alpha level X1000) of all statistics with alpha=.05 and delta=0.0 55

22 Size (Empirical alpha level X1000) of all statistics with alpha=.01 and delta=0.0 56

23 Power (X1000) of all statistics with alpha=.05 and delta=0.4 60

24 Power (X1000) of all statistics with alpha=.01 and delta=0.4 61

25 Power (X1000) of all statistics with alpha=.05 and delta=0.8 62

26 Power (X1000) of all statistics with alpha=.01 and delta=0.8 63

27 Statistics with significant power gains over over the paired t (in decreasing order) 75

I. INTRODUCTION

The one sample repeated measures design testing the equality of two correlated means (pre-test post-test paradigm) is commonly used in the evaluation of an educational program or treatment. It is an easy design to use and understand, and can be statistically more powerful than the use of two independent samples. Unfortunately, data collection exigencies may result in some missing observations for either or both time periods: subjects may be absent due to illness during testing; in schools with high student turnover rates, some students may move in shortly after the pre-test or out shortly before the post-test; poor testing conditions or instrument malfunction may invalidate responses or scores during one testing period, but not the other. In such cases the researcher must delete subjects from the analysis, thus reducing sample size.

Over the last ten years, a number of statisticians have begun to develop statistics which attempt to utilize both the paired and unpaired data which are available in repeated measures designs with missing observations. Initially, Lin(1971,1973), Mehta and Gurland(1969a,1969b,1973), Morrison(1973), and Naik(1975), studied the case with observations missing from one time period only. Each of these authors report that under certain circumstances their statistic is more powerful than a paired t-test conducted only on the data for subjects present at both time periods.

Unfortunately, none of these statistics appears to be consistently more powerful than the paired t under varying combinations of values for related parameters such as R - the ratio of variances (time 1/time 2), rho - the population correlation between time 1 and time 2, delta - the true difference between the means of time 1 and time 2, and L - the ratio of paired observations to all available observations. Lin's statistic, for example, is more powerful than the paired t only when the total number of observations is small and when rho is less than .5. Morrison's statistic performs well as the number of observations increases and when rho is equal to or greater than .3. Mehta and Gurland's statistic performs well when delta is large, and Naik's statistic is particularly useful when one variance is smaller than the other.

Over and above this is the consideration that all these statistics have been developed for data with observations missing from one time period only. Missing observations can occur at either testing time and so an adequate solution will take into consideration missing data at both times.

Lin and Stivers(1974), Ekbohm(1976), and Bohj(1978) more recently have begun to develop statistics which make use of missing observations from both time periods in a repeated measures design. Lin and Stivers' statistic ZLS and Ekbohm's statistic ZE, both based on maximum likelihood estimates of delta, were generally more powerful than other statistics developed by the same authors but based on a simple mean difference estimate of delta. Bohj's statistics are linear combinations of the paired and unpaired t. His statistics proved to have smaller values of the expected squared confidence interval length than the paired t. Bohj did not empirically examine the power of his statistics in comparison with the power of the paired t or any of the statistics developed by Lin and Stivers or by Ekbohm.

The purpose of this study, therefore, is to examine the statistics developed by Lin and Stivers, Ekbohm, and Bohj for the case with missing observations on both variables. The poewr and robustness of these statistics are compared with each other and with the paired t test for values of rho, delta and sample size which are relevant to the pre-test post-test educational paradigm.

A. NOTATION

Let X and Y be two normally distributed variables with means μ_X and μ_Y, variances σ_X and σ_Y, and correlation ρ. Let there be n pairs of observations on X and Y. In addition, let there be j unpaired observations on X, and k unpaired observations on Y (where applicable). Also let

$\bar{X}_p = \frac{1}{n} \sum_{i=1}^{n} X_p$ be the mean of the paired observations on X,

$\bar{Y}_p = \frac{1}{n} \sum_{i=1}^{n} Y_p$ be the mean of the paired observations on Y,

$\bar{X}_{up} = \frac{1}{j} \sum_{i=n+1}^{n+j} X_i$ be the mean of the unpaired observations on X,

$\bar{Y}_{up} = \frac{1}{k} \sum_{i=n+j}^{n+j+k} Y_i$ be the mean of the unpaired observations on Y,

$\bar{X}_a = \frac{1}{n+j} \sum_{i=1}^{n+j} X_i$ be the mean of all available observations on X,

$\bar{Y}_a = \frac{1}{n+k} \sum_{i=1}^{n} \sum_{i=n+j}^{n+j+k} Y_i$ be the mean of all available observations on Y,

$a_X = \sum_{i=1}^{n} (X_i - \bar{X}_p)^2$ be the sums of squares of the paired observations on X,

$a_Y = \sum_{i=1}^{n} (Y_i - \bar{Y}_p)^2$ be the sums of squares of the paired observations on Y,

$b_X = \sum_{i=n+1}^{n+j} (X_i - \bar{X}_{up})^2$ be the sums of squares of the unpaired observations on X,

$b_Y = \sum_{i=n+j}^{n+j+k} (Y_i - \bar{Y}_{up})^2$ be the sums of squares of the unpaired observations on Y,

$c_X = \sum_{i=1}^{n+j} (X_i - \bar{X}_a)^2$ be the sums of squares of all available observations on X,

$a_{XY} = \sum_{i=1}^{n} (X_i - \bar{X}_p)(Y_i - \bar{Y}_p)$ be the cross product of paired observations on X and Y,

$r = \dfrac{a_{XY}}{\sqrt{a_X \, a_Y}}$ and $u = \dfrac{2a_{XY}}{(a_X + a_Y)}$ be maximum likelihood estimates of the correlation coefficient rho,

$L = \dfrac{n}{n+j}$ be the ratio of paired observations to all observations,

$R = \dfrac{\sigma_X^2}{\sigma_Y^2}$ be the ratio of population variances, and

$\delta = \mu_X - \mu_Y$ be the true mean difference delta.

II. LITERATURE REVIEW

A. STATISTICS FOR DATA WITH MISSING OBSERVATIONS ON ONE VARIABLE

MEHTA AND GURLAND

Mehta and Gurland(1969a) proposed an estimate of delta (the true mean difference)

$$\delta = Z = \bar{Y}_p - A\,\bar{X}_p - (1-A)\,\bar{X}_{up}$$

where $A = L+u(1-L)$, $u = 2a_{XY}/(a_X + a_Y)$ and $L = n/(n+j)$.

They studied the size of its variance in comparison to the size of the variance of the simple mean difference $T = \bar{X}a - \bar{Y}p$. In this case u is a maximum likelihood estimate of rho when R=1 ($\sigma_X^2 = \sigma_Y^2 = \sigma^2$). They studied the efficiency (defined as var T/var Z) of the variance of Z when the absolute value of rho was .9, .7, .5, .3, and .1, for several values of n, for R=.1, .2, .5, 1.0, 2.0, 5.0, and 10.0, and when L=.1, .5, and .9(L is the ratio of paired to total observations, ie. n/(n+j)). They found that the estimator Z had a smaller standard error and was therefore a more efficient estimator of delta than the estimator T when the values of rho were large, and as the number of missing values increased. They also found that Z became increasingly more efficient as R became closer to 1. The variance of Z was not more efficient than the variance of T for decreasing values of R combined with decreasing values of the absolute value of rho, and for any absolute value of rho less than .3.

The calculation of the variance in this study was fairly cumbersome. Mehta and Gurland also admitted that the distribution of a T statistic using Z with A=L+u(1-L) was too complicated to be of practical use. In later articles (Mehta and Gurland 1969b,1973) they modified their work and used the statistic

$$\mathrm{TMG} = \frac{(\overline{Y}_p - B\overline{X}_p - (1 - B)\overline{X}_{up})^2}{A1a_X + A2a_Y + A3b_X + A4a_{XY}}$$

where the constants B, A1, A2, A3, and A4 were chosen so that a) the statistic had size alpha equal to or less than .05 for all values of rho greater than 0, and b) the gain in power over a paired t was optimized for all values of rho greater than zero. They provided tables for these constants for given n and j.

Mehta and Gurland presented data comparing the size and power of TMG in comparison to the paired t, for the following combinations of parameters: 1) positive values of rho: .1, .2, .3, .7, .9, 2) true difference delta of 0.0, .4, 1.5, 4.0 (no metric given for delta), and 3) n=3 to 17 and n+j=10 to 20. Results show that TMG was more powerful than the paired t for most of the values of rho that the authors tested. The gain in power over t increased as delta increased and as the proportion of missing data increased. The gain in power decreased as the total number of observations increased (n+j=20) and as rho increased (t was generally more powerful for rho=.9).

Mehta and Gurland provide tables of values for the constants required in their formula, for limited values of rho, delta and sample size. Use of their statistic has been restricted so far by the availability of these tables. Mehta and Gurland did not report on the use of their statistic under conditions of negative correlation and unequal variance.

MORRISON

Morrison(1973) presented a statistic based on maximum likelihood estimates of means and variances developed in an earlier paper(Morrison 1971). His estimate of the mean difference included the mean difference of all available observations on X and Y, adjusted by u, an estimate of rho based on the complete pairs of observations.

$$TM = \frac{\left\{\bar{X}_a - \bar{Y}_p - \delta + u(\bar{X}_p - \bar{Y}_{up})\frac{j}{n+j}\right\}\sqrt{(n+j)j(n+2k-3)}}{\sqrt{(a_X + a_Y + b_X)(n+2j+nu)(1-u)}}$$

This statistic is distributed as t with n-1 degrees of freedom. Morrison's statistic is limited to situations where the population correlation coefficient is equal to or greater than 0.0, and the variances of X and Y are unknown but equal.

Morrison evaluated his statistic by comparing the expected squared length of the confidence interval of TM with the expected squared length of the confidence interval of a paired t calculated on the n pairs of observations. He reported results for a limited variety of conditions: rho=0, .5, and .75; n=10 to 30; and j=5 to 50.

Although the TM statistic showed a smaller confidence interval for all these conditions, the gain was very slight if rho was large. The efficiency of TM increased as rho decreased, and as the proportion of missing observations increased. Thus this statistic is of little use for small samples and large values of rho.

NAIK

Naik(1975) proposed a statistic which uses constants ($\ell 1$ and $\ell 2$) to control the size to the preassigned alpha level for negative rho and all values of the two variances, as well as some positive values of rho when the variance of X is smaller than the variance of Y. ((He provided equations for $\ell 1$ and $\ell 2$, refering the reader to two articles on quadratic forms for their solution. His statistic is:

$$TN = \frac{\bar{Y}_p - L\bar{X}_p - (1 - L)\bar{X}_{up} - \delta}{\ell 1 \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \{X_i - \bar{X}_p - L(Y_i - \bar{Y}_p))\}^2 \right\} + \ell 2 \left\{ \frac{1}{j(j-1)} (1-L)^2 \sum_{i=n+1}^{j} (Y_i - \bar{Y}_{up})^2 \right\}}$$

Naik claims the variance of the numerator will be minimized by using L=n/(n+j) when the values of rho and R are unknown. Values of TN greater than 1 are significant. He compared the size and power of TN with the paired t for combinations of rho equal to 0, -.2, -.5, -.8, .1, .3, and .7; R equal to 1, 4, and .25; sample size up to 24; and (Delta/varX) equal to 0, .4, .9 and 1.5. Naik did not give the metric of delta. If the population variance equals 1.0, delta would be equal to 0.0, .63, .95, and 1.22.

Naik's statistic was more powerful than the paired t under many circumstances. Its superiority increased as R increased, that is, as the variance of the variable with missing observations increased. It appeared to be particularly applicable for negative and some positive but small values of rho. TN was not especially applicable if rho was positive or if the variance of the variable with the missing observations was smaller than the variance of the second variable, particularly for a small proportion of unpaired to paired observations.

LIN AND STIVERS

Lin(1973) used a maximum likelihood estimate of rho, $u = 2a_{XY}/(a_X + a_Y)$, in his statistic

$$TLS = \frac{\overline{X}_a - \overline{Y}_p - \delta}{\sqrt{\left\{\frac{L^2 - 2Lu + 1}{n} + \frac{(1 - L)^2}{j}\right\} \frac{a_Y - b_X}{n + j - 2}}}$$

He showed, by way of Monte Carlo studies that this statistic is approximately distributed as t with n-3 degrees of freedom for n equal to or greater than 5. He then compared TLS to the normal paired t under varying conditions of delta(0.0, 1.0, 3.0, and 5.0), and rho(.9, .5, .1, and 0.0). Lin's two populations were normally distributed with means equal to 0.0 and variances equal to 5.0. The values of delta therefore fell within the range of the population variances. The TLS statistic proved to be more powerful than the paired t except for high values of rho. For fixed rho, power increased with delta, and for fixed delta, power increased as the value of rho increased. Lin did not present data with regard to the performance of TLS with unequal variances on X and Y.

Using Monte Carlo methods, Lin and Stivers(1975) compared four statistics (TMG, TM, TLS, and the paired t) under conditions of equal variances, rho= .1 to .9, n of 5 to 30, n+j of 10 to 90, and delta of 0.0 to 2.0. In this case, the two population means were 0.0 and the variances were 1.0. Values of delta went beyond the value of the population variances. They found that no test was uniformly more powerful for all values of rho.

Table 1 shows the parameters studied by Lin and Stivers and the one most powerful statistic for each condition. The paired t was the most powerful statistic when rho was equal to or greater than .9. TLS was the most powerful statistic for small values of rho (and some higher values of rho if the number of observations was less than 15). TMG was rarely

TABLE 1

The most powerful statistic(s)
for each condition of rho, delta, n, and j -
from Lin and Stivers(1975)

rho	n=7	n=10	n=20	n=20	n=30	n=30
	j=3	j=5	j=20	j=30	j=30	j=60
-.9	TLS*	TMG	TM	TM	TM	TM
	TLS	TLS	TM	TM	TM	TM
-.7	TLS	TLS	TM	TM	TM	TM
	TLS	TLS	TM	TM	TM	TM
-.5	TLS	TLS	TM	TLS	TM	TM
	TLS	TLS	TM	TM	TM	TM
-.3	TLS	TLS	TLS	TLS	TLS	TLS
	TLS	TLS	TLS	TM	TM	TM
-.1	TLS	TLS	TLS	TLS	TLS	TLS
	TLS	TLS	TLS	TM	TLS	TM,TLS
.1	TLS	TLS	TLS	TLS	TLS	TM
	TLS	TLS	TLS	TLS	TLS	TM,TLS
.3	TMG	TLS	TLS	TLS	TM	TLS,TM
	TMG	TMG	TM	TM	TM	TM
.5	TM	TMG	TM	TM	TM	TM
	TM	TMG	TM	TM	ALL	ALL
.7	TM	TM	TM	TM	TM	TM
	TM	TM	TM	t	ALL	ALL
.9	t	t	t	t	TM,t	TM,t
	t	t	t	t	ALL	ALL

TLS - Lin and Stivers
TM - Morrison
TMG - Mehta and Gurland
t - paired t

*Results are for delta=.5 and 1.0

more powerful than the others. In a few instances TMG was more powerful for small values of delta. TM was the most powerful statistic when n was moderately large and rho was between .3 and .9. This statistic performed the best as the number of observations increased.

EKBOHM

Ekbohm(1976) studied the same four statistics (TLS, TM, TMG, and the paired t) under similar conditions to those of Lin and Stivers(1975) but restricted himself to positive values of rho and small n's. With population means of 0.0 and variances of 1.0, his values of delta equal to .3, .6, and 1.0 are within the range of the population variances. Table 2 shows the results of his study. TLS was the most powerful statistic for small n and medium to small rho. TM was generally the most powerful statistic for higher values of rho and n. The paired t was the most powerful statistic for values of rho equal to or greater than .9. These results, for the most part, coincide with those of Lin and Stivers.

Ekbohm also studied the same four statistics under conditions of unequal variances. Table 3 shows these results. TLS was the most powerful statistic if the variable with the missing observations had the larger variance. TM was the most powerful statistic if the variable with the missing observations had the smaller variance. Unfortunately, all these statistics, except the paired t,

TABLE 2

The most powerful statistic(s)
for conditions of positive rho and equal variances -
from Ekbohm(1976)

rho	delta	n=10,j=2	n=10,j=5	n=10,j=10
0.0	.3	TLS	TLS	TLS
	.6	TLS	TLS	TLS
	1.0	TM,TMG	TLS	TLS
.2	.3	TLS	TLS	TLS
	.6	TLS	TLS	TLS
	1.0	TMG	TLS	TLS
.5	.3	TLS	TLS	TM
	.6	TLS	TMG	TMG
	1.0	TM	TMG	TMG
.8	.3	TM	TM	TM
	.6	TM	TM	TM
	1.0	TM	TM	t,TM,TMG

TABLE 3

The most powerful statistic
for conditions of positive rho and unequal variances
when n=10 and j=5 - from Ekbohm(1976)

rho	delta	varX=2varY	varX=.5varY
0.0	.3	TLS	TM
	.6	TLS	TM
	1.0	TLS	TM
.2	.3	TLS	TM
	.6	TLS	TM
	1.0	TLS	TM
.5	.3	TLS	TM
	.6	TLS	TM
	1.0	TLS	TM
.8	.3	TLS	TM
	.6	TM	TM
	1.0	TMG	TM

TLS - Lin and Stivers
TMG - Mehta and Gurland
TM - Morrison
t - paired t

had too large a size under many conditions, when the variances of X and Y were not equal.

SUMMARY OF PROBLEMS

The authors cited above were all attempting to design a statistic which, by making use of additional data from unpaired observations resulting from missing observations in a pre-test post-test paradigm, would be more powerful than the paired t under varying conditions of rho, delta, R and sample size. There are, however several problems with these statistics.

1. Each statistic, under one or more sets of conditions is less powerful than the paired t. For example, TLS and TMG are less powerful than the paired t when rho is high (.7 to .9). TM is less powerful than the paired t when the variances of X and Y are unequal and when the true difference delta is small. Thus one cannot use one statistic exclusively and expect to gain power over the paired t under all circumstances.
2. When the variances of X and Y are unequal, the choice of the most powerful statistic is dependent on whether the variable with the missing observations has a larger or smaller variance. TLS is more powerful when the variable with the missing observations has a larger variance, TM is more powerful when the variable with the missing observations has the smaller variance. Again, no one statistic can be used for maximal advantage in both situations.
3. The gain in power for all statistics over the paired t decreases as rho increases.

4. All these statistics are designed for missing observations on one variable only. In many situations, observations are missing from both X and Y variables.

B. STATISTICS FOR DATA WITH MISSING OBSERVATIONS ON BOTH VARIABLES

LIN AND STIVERS

Lin and Stivers(1974) were the first authors to develop a T statistic to be used when both X and Y have missing observations. They developed a maximum likelihood estimate of delta, using estimates of the population correlation coefficient and variance based on the paired data. They proposed the following statistic using this estimate of delta:

$$ZLS = \frac{\delta^* - \delta}{\hat{\gamma}}$$

where $\delta^* = \hat{a}\overline{X}_p + (1 - \hat{a})\overline{X}_{up} - \hat{b}\overline{Y}_p - (1 - \hat{b})\overline{Y}_{up}$,

$$\hat{a} = n\hat{h}(n + k + jv) , \quad \hat{b} = n\hat{h}(n + j + kw) ,$$

$$\hat{h} = \frac{1}{(n + j)(n + k) - jkr^2} ,$$

$$v = \frac{a_{XY}}{a_X} , \quad w = \frac{a_{XY}}{a_Y} , \quad r = \frac{a_{XY}}{\sqrt{a_X a_Y}} , \text{ and}$$

$$\hat{\gamma}^2 = \left\{\frac{\hat{a}^2}{n} + \frac{(1-\hat{a})^2}{j}\right\}\frac{a_X}{n - 1} - \frac{2\hat{a}\hat{b}}{n}\,\frac{a_{XY}}{n - 1} + \left\{\frac{\hat{b}^2}{n} + \frac{(1-\hat{b})^2}{k}\right\}\frac{a_Y}{n - 1}$$

This statistic is distributed as t with n degrees of freedom.

Lin and Stivers proposed two additional but approximate statistics to test the hypothesis of no difference between means. These statistics are based on the simple difference of sample means, based on all the available data.

1. TLS1 was designed for conditions where R=1 and rho is unknown. This statistic is distributed as t with n+j+k-4 degrees of freedom.

$$TLS1 = \frac{\overline{X}_a - \overline{Y}_a - \delta}{\sqrt{\frac{1}{n + j} + \frac{1}{n + k} - \frac{2nr}{(n + j)(n + k)}}\sqrt{\frac{c_X b_Y}{n + j + k - 2}}}$$

2. TLS2 was based on Welch's(1947) approximation to the Behrens Fisher problem in the case of unequal variances.

$$TLS2 = \frac{\overline{X}_a - \overline{Y}_a - \delta}{h_1 + h_2 + h_3}$$

$$\text{where} \quad h_1 = \frac{\frac{n(n + k)a_X}{n + j} + \left\{\frac{(n + j)a_Y}{n + k} - 2a_{XY}\right\}}{(n - 1)(n + j)(n + k)},$$

$$h_2 = \frac{jb_X}{(j - 1)(n + j)^2}, \text{ and } h_3 = \frac{kb_Y}{(k - 1)(n + k)^2} .$$

This statistic is distributed as t with degrees of freedom given as

$$df = \frac{(h_1 + h_2 + h_3)^2}{\frac{h_1^2}{n - 1} + \frac{h_2^2}{j - 1} + \frac{h_3^2}{k - 1}}$$

Lin and Stivers compared these three statistics, ZLS, TLS1, and TLS2 with the paired t with regard to size and power, in 1000 simulated random samples from a bivariate normal distributions, under each of the following conditions:

1. rho=-.9, -.5, -.1, 0, .1, .5, and .9,
2. R=.25, .5, 1, 2, and 4,
3. n=5 to 20, j=5 and 9, k=3 and 10, and
4. delta=0.0, 1.0, and 3.0 (no metric given for delta).

A statistic was considered satisfactory with regard to size if the observed frequency of the test at delta=0.0 did not exceed the nominal alpha level by more than $\pm 2\sqrt{\alpha(1 - \alpha)/1000}$. The ZLS statistic met this criterion under most conditions when n was greater than 10. TLS1 met the criterion when n was between 5 and 20, provided R was equal to or near unity. TLS2 met the criterion when n was less than 20, regardless of the value of R.

With regard to power, Lin and Stivers found that for all four statistics, (TLS1, TLS2, ZLS, and paired t) for fixed R, the power increased with rho, and for fixed rho, the power decreased as R deviated from unity. The ZLS statistic was the most powerful statistic under conditions of moderate correlation and n larger than 15. When n was between 5 and 15, TLS1 and TLS2 were most powerful for lower values of rho and R near unity. All the statistics proposed by Lin and Stivers were more powerful than the paired t except when rho was equal to .9. Tables 4, 5, 6 and 7 show Lin and Stivers' evaluation of the one best statistic (based on the results of both the size and power studies) for varying conditions of rho, delta, R, n, j, and k. These tables show that ZLS was generally the best statistic when the number of observations was large and rho was equal to or greater than .5. For smaller n's and lower correlations, TLS1 performed best when R was close to unity. TLS2 was more sensitive to deviations from equal variance assumptions. Results in Table 5 suggest that TLS2 however may be sensitive to the number of missing observations on the variable with the larger variance.

EKBOHM

Ekbohm(1976) used Monte Carlo methods to evaluate the size and power of five statistics for testing the equality of means in the case of missing observations on both variables. Three of these were from Lin and Stivers' study (ZLS, TLS1, and TLS2). Two others he proposed himself. The first, ZE, used a maximum likelihood estimate of delta similar to that of Lin and Stivers, substituting u for $\hat{a}$ and $\hat{b}$. (See page 17 for definitions of $\hat{a}$ and $\hat{b}$.) This estimate

TABLE 4

The preferred statistic(s)
for varying conditions of rho and R(varX/varY)
when n=5, j=5, k=10, and delta=1 and 3 -
from Lin and Stivers(1974)

rho	R=.25	R=.5	R=1.0	R=2.0	R=4.0
-.9	TLS2	*	*	*	TLS2
-.5	TLS2	TLS1&2	TLS1	TLS1	TLS2
-.1	TLS2	TLS1&2	TLS1	TLS1	TLS2
0.0	TLS2	TLS1&2	TLS1	TLS1	TLS2
.1	TLS2	TLS1&2	TLS1	TLS1	TLS2
.5	TLS2	TLS1&2	TLS1	TLS1	TLS2
.9	TLS2	t	t	t	t

TABLE 5

The preferred statistic(s)
for varying conditions of rho and R(varX/varY)
when n=10, j=5, k=10, and delta=1 and 3 -
from Lin and Stivers(1974)

rho	R=.25	R=.5	R=1.0	R=2.0	R=4.0
-.9	TLS1	TLS1	TLS1	TLS1	*
-.5	TLS1	TLS1	TLS1	TLS1	*
-.1	TLS1	TLS1	TLS1	TLS1	TLS2
0.0	TLS1	TLS1	TLS1	TLS1	TLS2
.1	TLS1	TLS1	TLS1	TLS1&2	TLS2
.5	TLS1	TLS1	TLS1&2	TLS1	TLS2
.9	ZLS	ZLS	ZLS	ZLS	ZLS

*Lin and Stivers do not specify the preferred statistic(s) for these conditions.

TABLE 6

The preferred statistic(s)
for varying conditions of rho and R(varX/varY)
when n=15, j=5, k=10, and delta=1 and 3 -
from Lin and Stivers(1974)

rho	R=.25	R=.5	R=1.0	R=2.0	R=4.0
-.9	ZLS	ZLS	ZLS	ZLS&TLS2	ZLS&TLS2
-.5	ZLS&TLS2	ZLS&TLS1	ZLS&TLS1	ZLS&TLS2	ZLS&TLS2
-.1	TLS2	TLS1	TLS1	TLS2	TLS2
0.0	TLS2	TLS1	TLS1	TLS2	TLS2
.1	TLS2	TLS1	TLS1	TLS2	TLS2
.5	ZLS&TLS2	ZLS&TLS1	ZLS&TLS1	ZLS	ZLS
.9	ZLS	ZLS	ZLS	ZLS	ZLS

TABLE 7

The preferred statistic(s)
for varying conditions of rho and R(varX/varY)
when n=20, j=5, k=10, and delta=1 and 3 -
from Lin and Stivers(1974)

rho	R=.25	R=.5	R=1.0	R=2.0	R=4.0
-.9	ZLS	ZLS	ZLS	ZLS	ZLS
-.5	ZLS	ZLS	ZLS&TLS2	ZLS	ZLS
-.1	ZLS	ZLS	TLS1	ZLS&TLS2	ZLS
0.0	ZLS	ZLS	TLS1	ZLS&TLS2	ZLS
.1	ZLS	ZLS	TLS1	ZLS&TLS2	ZLS
.5	ZLS	ZLS	ZLS&TLS1	ZLS	ZLS
.9	ZLS	ZLS	ZLS&TLS1	ZLS	ZLS

is divided by an estimate of its standard error which uses a derived common variance estimator θ^2. ZE is distributed as t with n degrees of freedom:

$$ZE = \frac{\delta^* - \delta}{\sqrt{\theta^2 \, \dfrac{2n(1-u) + (j+k)(1-u^2)}{(n+j)(n+k) - jku^2}}}$$

$$\text{where} \quad \theta^2 = \frac{\sum\limits_{i=X,Y} a_i + (1+u^2) \sum\limits_{i=X,Y} b_i}{2(n-1) + (1-u^2)(j+k-2)}$$

Similarly, assuming equal variances on X and Y, Ekbohm derived an estimate of the variance of the simple mean difference ($\overline{X} - \overline{Y}$) and used this in the statistic SE:

$$SE = \frac{\overline{X}_a - \overline{Y}_a \sqrt{(n+j)(n+k)}}{\sqrt{m+d}}$$

$$\text{where} \quad m = \frac{n\left(\sum\limits_{i=X,Y} a_i - 2a_{XY}\right)}{n-1} \quad \text{and} \quad d = \frac{(j+k)\sum\limits_{i=X,Y} b_i}{j+k-2}$$

Ekbohm claims SE is distributed as t with degrees of freedom

$$df = \frac{(m+d)^2}{\dfrac{m^2}{n+1} + \dfrac{d^2}{j+k}} - 2$$

Table 8 summarizes the characteristics of these five statistics.

Ekbohm reported results comparing these five statistics with both paired and unpaired t statistics for values of

TABLE 8

Characteristics of statistics

Statistic	Scedasticity assumption	type of estimate
ZLS	hetero	modified MLE
TLS2	hetero	mean diff
ZE	homo	modified MLE
SE	homo	mean diff
TLS1	homo	mean diff

rho=0.0, .5, and .8, R=.5, 1.0, and 2.0, n=7 and 10, j=5 and 10, k=2 and 5, and delta=0.0, .6, and 1.0. (No metric was given for delta.) Results for each statistic were as follows:

1. ZLS proved to have too large a size for small values of n, as Lin and Stivers had reported. (Size was considered too large if the observed frequencies in 1000 samples with delta equal to zero surpassed the .05 alpha level by more than $\pm 2\sqrt{\alpha(1-\alpha)/1000}$.) Ekbohm found that the critical size of n was affected by the size of the difference between j and k, if the variances were unequal. In cases where the size of ZLS was acceptable, it was more powerful than the paired t.
2. ZE was as powerful as ZLS for all values of n, j, and k. Although ZE was designed assuming homoscedasticity, it appeared to be fairly robust with regard to departures from this assumption. ZE was too large however when j and k were not equal, the larger sample having the smaller variance.

3. SE and TLS1 were approximately equal in power. SE was more robust for unequal variances than was TLS1 when j=k. Neither statistic was more powerful than the paired t when rho was equal to or greater than .8. TLS1 was too large in size if the variance of X was twice that of Y.
4. TLS2 was generally more powerful than the paired t. It was also more robust than TLS1 when the variances of X and Y were unequal, especially if j and k were unequal. Like SE and TLS1, TLS2 was not more powerful than the paired t when rho was equal to or greater than .8.

In summary, the mean difference estimators TLS1, TLS2, and SE proved more powerful than the paired t only for low correlations (rho equal to or less than .5). The statistics based on maximum likelihood estimates, ZE and ZLS, were more powerful than the paired t for all values of rho. However, ZE and in particular ZLS were frequently too large in size. The sizes and ratios of n, j, and k had a deleterious affect on the size and power of ZE, SE, and TLS1 (all based on the homoscedastic assumption) such that no one statistic was applicable in all the situations examined. ZE seemed to be applicable in the widest variety of situations, although it was too large in some cases. No statistic was uniformly better than the paired t when n was less than 10. Ekbohm did not examine the performance of any of these statistics under the condition of negative values of rho.

BOHJ

Both Lin and Stivers(1974) and Ekbohm(1976) approached the problem of testing the hypothesis of no difference between means with missing observations on both variables by deriving an estimate of the mean difference and its variance. Bohj(1978) approached the same problem by using a linear combination of the paired and unpaired t. He proposed two statistics: TB1 for cases when R=1,

$$TB1 = q\frac{\overline{X}_p - \overline{Y}_p - \delta}{\frac{s_p}{\sqrt{n}}} + (1 - q)\frac{\overline{X}_{up} - \overline{Y}_{up} - \delta}{s_{up}\sqrt{\frac{1}{j} + \frac{1}{k}}}$$

$$\text{where} \quad s_p^2 = \frac{\sum_{i=1}^{n} (X_i - Y_i - \overline{X}_p - \overline{Y}_p)^2}{n - 1},$$

$$s_{up}^2 = \frac{\sum_{i=1}^{j} (X_{n+i} - \overline{X}_{up}) + \sum_{i=1}^{k} (Y_{n+i} - \overline{Y}_{up})^2}{j + k - 2}, \text{ and}$$

q is a constant of Bohj's choice $(0.0 \le q \le 1.0)$

and TB2 for unequal variances

$$TB2 = q\frac{\overline{X}_p - \overline{Y}_p - \delta}{\frac{s}{\sqrt{n}}} + (1 - q)\frac{\overline{X}_{up} - \overline{Y}_{up} - \delta}{\frac{s_2}{\sqrt{j}}}$$

$$\text{where} \quad s_2^2 = \frac{\sum_{i=1}^{j} (w_i - \overline{w}_X)}{j - 1},$$

$$w_i = X_{n+i} - \sqrt{\frac{j}{k}}\, Y_{n+j}, \text{ and } \overline{w}_X = \sum_{i=1}^{j} \frac{w_i}{j}$$

Following the method of Patil(1965) for the distribution of weighted sums of two independent variables distributed as t, Bohj claims that his statistic multiplied by a constant h, is distributed as t with f degrees of freedom. H and f are derived by Patil as follows:

$$f = 4 + \frac{\left. \dfrac{s_{up}^{2} f_2}{f_2 - 2} + \dfrac{s_p^{2} f_1}{f_1 - 2} \right.^{2}}{\dfrac{s_{up}^{4} f_2^{2}}{(f_2 - 2)^2 (f_2 - 4)} + \dfrac{s_p^{4} f_1^{2}}{(f_1 - 2)^2 \, (f_1 - 4)}}$$

where f_1 are the degrees of freedom for the paired t

and f_2 are the degrees of freedom for the unpaired t

$$h^2 = \frac{\dfrac{f}{f - 2}}{\sqrt{\dfrac{s_{up}^{2} f_2}{f_2 - 2} + \dfrac{s_p^{2} f_1}{f_1 - 2}}}$$

Bohj evaluated his statistics by comparing the expected squared length of the confidence intervals of TB1 and TB2 $(E(L^2_{TB1}))$ and $(E(L^2_{TB2}))$ with that of the paired t $(E(L^2_t))$ With $C=E(L^2_{TB1})/E(L^2_t)$, his value for q was chosen to minimize the average value of C for all values of rho examined for a given n, j, and R. He also used q=.5, which resulted in an unweighted linear combination of the paired and unpaired t.

For TB1 Bohj examined conditions of rho=.1, .3, .5, .7, and .9, n=10, j=5 to 30, and k=5 to 30. His results showed that the confidence interval of TB1 was shorter than that of the paired t for all reported conditions when the weighted q was used. When an unweighted TB1 was calculated , $(E(L^2_{TB1}))$ was greater than $(E(L^2_t))$ for small values of j and k when rho was .9. Overall C increased as rho increased and as j and k increased. Thus, TB1 showed the greatest gains over the paired t when the correlation between X and Y was small, and when the number of missing observations was large, relative to the number of paired observations.

Using TB2, (for unequal variances), Bohj examined conditions of rho=.1, .3, .5, .7, and .9, R=.6 and 4.0, n=10, j=6 and 10, and k=10 and 30. In this case, the value of C was greater than 1 under one condition only: rho=.9, R=.6,j=6 and k=10. As was the case with TB1, the value of C decreased as rho decreased and as the number of missing observations increased. TB2 was more efficient for values of R=.6 than for R=4.0. This effect became more pronounced as rho increased. The gain in precision of TB2 over the paired t increased as the number of observations on the variable with the largest variance increased.

For both TB1 and TB2 the gain in precision of the weighted statistic over the unweighted statistic appeared to be primarily a function of Bohj's method of choosing q to minimize the average value of C for all values of rho for a given number of observations. As a result, the weighted statistic was more precise (that is, it resulted in a lower value of C) than the unweighted statistic for high values of rho and for small values of j and k. Weighting had little effect on decreasing values of rho combined with increasing j and k.

In conclusion, the expected lengths squared of the confidence intervals of Bohj's statistics TB1 and TB2 were generally smaller than those of the paired t when used in evaluating the signifigance of the difference between correlated means for various values of rho and sample size. Bohj did not test his statistics via a Monte Carlo type study, and therefore could not empirically evaluate the size, power, or fit of their t distributions.

III. METHODOLOGY

A. INTRODUCTION

This study examined the performance of seven statistics which have been developed to test the equality of means in the one sample repeated measures design with missing values on both measures. Six new statistics : ZLS, TLS1, and TLS2 developed by Lin and Stivers; ZE and SE developed by Ekbohm; and TB1 developed by Bohj were compared to the paired t.

Each of these statistics was examined under the nine combinations of rho equal to .5, .7, .9, and delta equal to 0.0, .4, and .8. Only high values of rho were used as these are most common in educational pre-test post-test research designs. The choice of values for delta was based on the power results from the one variable studies which gave the metric they used for delta (Naik(1975), Lin and Stivers(1973,1975) and Ekbohm(1976)). In these studies, power values reached the ceiling value of 1.00 when the value of delta was equal to or greater than that of the population variances. For each of the nine combinations of rho and delta, the following nine combinations of n, j, and k were examined:

n	8	8	8	16	16	16	24	24	24
j	8	8	16	8	8	16	8	8	16
k	8	16	16	8	16	16	8	16	16

B. PROCEDURES

CREATION OF POPULATIONS

A total of nine populations, each with 3000 observations, was required for the study; that is, one population for each of the nine combinations of rho and delta (rho =.5, .7, and .9; delta =0.0, 0.4, and 0.8). These populations were generated using a fortran program written by S. Hunka, of the University of Alberta, which is based on matrix transformation methods taken from Kaiser and Dickman (1972). Three pairs of variables were generated for each of the nine combinations. In each case, the one pair of variables showing the best fit to the normal distribution (that is, highest probability values on the Kolmogorov-Smirnov goodness of fit test) was used in the study.

SAMPLING AND CALCULATION OF STATISTICS

Using the data analysis program in Appendix I, 1000 random samples were selected (with replacement)[1] from each of the nine populations for each of the nine combinations of sample size outlined above. This produced a total of 9 X 9 =81 sample sets. In each case, the initial seed used was obtained from the subroutine documentation. Subsequent runs began with the last seed generated in the previous run.

---

[1] A subroutine from the International Mathematics and Statistical Libraries (see Bibliography) was used to create the required random normal deviates.

In order to ascertain the adequacy of the samples generated, the distributions of six means ($\bar{X}p$, $\bar{Y}p$, $\bar{X}up$, $\bar{Y}up$, $\bar{X}a$, and $\bar{Y}a$) were checked for each sample set using the Kolmogorov-Smirnov goodness of fit test. The means and standard deviations of these distributions were compared to the expected values of the means and the standard errors of the mean.

The program in Appendix I also calculated the value of each satatistic, its degrees of freedom and the associated probability level for each of the 1000 samples in each sample set, according to the formulae given in Lin and Stivers(1974), Ekbohm(1976) and Bohj(1978). Bohj's statistic TB1 was calculated using values of q equal to .1, .3, .5, .7, and .9, with a view to choosing the one best value, for comparison with the other statistics.

EVALUATION OF STATISTICS

The adequacy of each statistic was judged according to performance on three criteria:

1. Size: In those sample sets with delta=0.0, the empirical alpha level was calculated as the proportion of probability levels less than the nominal alpha levels of .05 and .01. The size of a statistic was judged to be 1)acceptable if this proportion was within two standard deviations of the nominal alpha levels, 2)fair if this proportion was within two to three standard deviations of the nominal alpha level, and 3)poor if this

proportion was beyond three standard deviations of the nominal alpha level.

2. Power: For those sample sets with delta greater than 0.0, the power of a statistic was calculated as the proportion of probability levels less than the nominal alpha levels of .05 and .01. Power was judged to be 1)good if this proportion was over .90, 2)fair if this proportion was between .7 and .9, and 3)poor if this proportion was below .70 (see Kendal and Stuart, 1967). By making use of information not used in the traditional paired t test, the developers of each of the statistics examined in this study have proposed statistics which they hope will be more powerful than the paired t. Therefore, each statistic was also evaluated according to if and when it was more powerful than the paired t for various values of rho, delta, and sample size.
3. Goodness of fit to the appropriate t distribution: The goodness of fit of each statistic to the appropriate t distribution was ascertained by probability levels from Kolmogorov-Smirnov goodness of fit tests. Chi Square goodness of fit tests were not used as they are subject to two problems: 1) combining small expected frequencies may violate the assumption that the observed values are normally distributed around the expected values, and 2) the choice of the number of categories can lead to different results. Siegal(1956) and Henkel(1976) feel that the Kolmogorov-Smirnov test is usually more

powerful than the Chi Square test, particularly for small sample sizes (as is the case with TB1, TLS2 and SE) and for continuous variables. Kendall(1967) feels that the Kolmogorov-Smirnov test is more efficient than the Chi Square test, as well as more powerful. The formulae used to calculate the degrees of freedom for the statistics TB1, TLS1 and SE, include some sample statistics. Therefore, when 1000 samples were created for a given sample set, a range of non-integer degrees of freedom resulted. For purposes of the goodness of fit tests, these degrees of freedom were rounded to the nearest integer value. A Kolmogorov-Smirnov goodness of fit test was then performed for each integer degree of freedom, in each sample set, which had an N of over 60. The adequacy of the fit of a statistic could only be judged roughly as the probability levels showed a large variation. Fit was judged to be acceptable if 1)the majority of probability values were within the same range as those of the paired t, and 2)very few probability values were less than .20.

IV. RESULTS

A. ADEQUACY OF DATA GENERATION PROCEDURES

POPULATION

The population generation program provided population data with the desired population parameters and with excellent fit to the normal distribution. Each variable had a variance of 1.0. All X variables had a mean of 0.0. The Y variables had means of 0.0, 0.4, or 0.8, depending on the value of delta. The correlations between variables were exactly .5, .7, and .9. Kolmogorov-Smirnov probability values for the goodness of fit to the normal distribution are given in Table 9 for each of the populations used in the study. The probability values are for each pair of variables (N=3000) chosen from the six generated for each of the nine combinations of rho and delta.

SAMPLE

The distributions of six means ($\bar{X}p$, $\bar{Y}p$, $\bar{X}up$, $\bar{Y}up$, $\bar{X}a$, and $\bar{Y}a$) were obtained for each sample set of 1000 observations. These distributions proved to be quite adequate for all 81 sample sets. In every case, the mean of the distribution was well within one standard error of the expected value of the mean. All standard deviations fell within plus or minus .03 points of the expected values of the standard errors of the mean. (Expected values of the

TABLE 9

Goodness of fit of populations to the Normal Distribution

rho	delta	K-S* p-values variable X	K-S* p-values variable Y
.9	0.0	1.0	1.0
	.4	1.0	1.0
	.8	1.0	.98
.7	0.0	1.0	1.0
	.4	1.0	1.0
	.8	1.0	1.0
.5	0.0	1.0	1.0
	.4	.94	1.0
	.8	1.0	1.0

*K-S: Kolmogorov-Smirnov

standard errors of the mean ranged from .134 to .354, depending on sample size). Kolmogorov-Smirnov probability values for goodness of fit to the normal distribution for these distributions ranged from .24 to 1.00, with the majority being above .70. Kolmogorov-Smirnov probability values for the goodness of fit of the paired t to the t distribution ranged from .39 to 1.00. (Complete figures for the fit of the paired t are given in Table 10 below.) The size (empirical alpha level) of the paired t was always within two standard errors of the nominal alpha level, the size standard used in studies by Lin and Stivers and by

Ekbohm. (Complete figures for the size of the paired t are given in Table 21.)

B. EVALUATION OF STATISTICS

GOODNESS OF FIT TO THE t DISTRIBUTION

The goodness of fit of each of the new statistics to the appropriate t distribution was compared to the goodness of fit of the paired t to its t distribution when delta was equal to 0.0. Table 10 gives the Kolmogorov-Smirnov goodness of fit probability values for the paired t, ZLS, TLS1, and ZE. The degrees of freedom for each of these statistics was invariant within a sample set. Probability values for TLS1 ranged from .29 to 1.0 while probability values for ZE ranged from .21 to 1.0. For both of these statistics the majority of probability values were greater than .70. These probability values were generally within the same range as those of the paired t. Probability values for ZLS tended to be quite a bit lower than those of the paired t when n (the number of paired observations) was low and the number of missing observations was high (that is, sample sets 8816 and 81616). Only three probability values were lower than .20, however. Probability values for ZLS when n was greater than eight were comparable to those of the paired t, that is, most were greater than .70.

The degrees of freedom for TB1, TLS2, and SE are variant since sample statistics are used in their

TABLE 10

Kolmogorov-Smirnov p-values for the goodness of fit of the paired t, ZLS, TLS1, and ZE to their repective t distributions (delta=0.0)

sample set	paired t	ZLS	TLS1	ZE
888	.95*	.98	1.0	1.0
	.80	.54	1.0	.64
	1.0	.55	1.0	.88
8816	1.0	.19	.78	.76
	1.0	.22	.74	.89
	.58	.01	1.0	.31
81616	.82	1.0	1.0	1.0
	.44	.36	.72	1.0
	.92	.22	.83	.99
1688	1.0	.85	.56	.99
	.68	.47	.87	.21
	.92	.61	.96	.67
16816	.39	.46	1.0	.69
	.71	.63	.97	1.0
	1.0	.71	1.0	.99
161616	1.0	1.0	.81	1.0
	.90	1.0	.70	1.0
	1.0	.99	.99	.80
2488	.75	.96	.59	.97
	.58	.97	1.0	.99
	.99	1.0	1.0	.92
24816	1.0	1.0	1.0	.97
	.95	.10	.29	.45
	.97	.96	.99	.93
241616	.95	.94	.68	.76
	1.0	.67	.45	.81
	1.0	.98	.59	1.0

*The three figures refer to rho=.5, .7, and .9 respectively calculation. Table 11 provides data on the range of degrees of freedom (that is, the number of different degrees of freedom) for each of these statistics. The range of degrees of freedom was relatively constant for varying values of delta for a given combination of rho and sample size for each statistic. The range of degrees of freedom varied somewhat for the three values of rho within a sample size. For TLS2, the number of degrees of freedom increased as rho increased; for TB1 and SE no particular trend was evident. The greatest change in range of degrees of freedom occurred across sample size. This change did not appear to be related to any of the conditions examined in this study. The data in Table 11 do indicate that for all three of these statistics, (and for TLS2 in particular, where over half the sample sets produced a range of over 20 different degrees of freedom) the range of degrees of freedom was very large. This meant that the number of samples or N for any one degree of freedom and overall, the N's for the goodness of fit tests for these three statistics were much smaller than the N of 1000 for the other four statistics.

Kolmogorov-Smirnov goodness of fit tests were calculated separately for each integer degree of freedom for TB1, TLS2, and SE. Tables 12 to 20 show the probability values for these goodness of fit tests. Since the number of samples for many degrees of freedom were extremely small, Tables 12 to 20 only provide data on the six degrees of freedom with the largest N, or those degrees of freedom

TABLE 11

Range of Degrees of freedom for TB1, TLS2, and SE

Sample set	delta	TB1			TLS2			SE		
		0.0	0.4	0.8	0.0	0.4	0.8	0.0	0.4	0.8
	rho									
888	.5	10	10	10	12	12	12	12	13	13
	.7	9	10	10	12	14	13	9	13	10
	.9	8	7	6	14	14	13	10	10	10
8816	.5	18	20	18	18	20	19	14	14	17
	.7	17	17	18	18	20	19	8	9	12
	.9	14	15	13	20	21	20	9	10	10
81616	.5	26	26	26	19	19	18	18	18	17
	.7	25	25	25	16	16	16	9	9	9
	.9	22	21	23	18	19	16	9	10	10
1688	.5	9	8	8	17	17	15	13	13	12
	.7	10	10	10	18	19	18	15	15	15
	.9	11	11	11	21	19	20	17	17	17
16816	.5	14	14	14	22	21	21	14	14	14
	.7	10	10	12	26	24	24	13	14	14
	.9	11	11	10	28	27	26	16	16	17
161616	.5	22	22	23	17	17	20	17	17	19
	.7	17	20	18	19	21	19	16	15	14
	.9	11	10	10	21	23	23	15	16	17
2488	.5	17	16	16	22	22	20	18	15	17
	.7	17	17	18	23	25	25	19	20	20
	.9	18	19	15	24	26	26	23	24	24
24816	.5	12	11	13	26	25	32	15	15	17
	.7	14	14	16	30	30	31	16	19	21
	.9	18	17	18	30	32	33	23	23	22
241616	.5	15	17	14	22	18	21	14	16	15
	.7	11	15	12	23	22	23	18	21	19
	.9	17	17	17	27	25	27	21	21	23

containing over sixty samples.

The probability values for the goodness of fit of TLS2 ranged from .04 to 1.0. They were generally within the same range as those of the paired t, with only a few tending to be slightly lower. Sample sets 8816, 81616, 16816 and 24816, with rho=.9 consistently had the highest probability values. Since the range of degrees of freedom for TLS2 was so large, the number of samples, or N, of most individual degrees of freedom was less than 60. In only a few instances was N greater than 200. This made it difficult to predict how adequate the fit would be for an N of 1000, as was the case with ZLS, TLS1, ZE and the paired t. However there was no indication in the data that the adequacy of the fit changed with increasing or decreasing N, within any of the sample sets.

The probability values for the goodness of fit of SE ranged from .00 to 1.0, with most being extremely high, and well within the range of the paired t, for all combinations of rho and sample size. The range of degrees of freedom for a given sample set was generally less than that of TLS2. As a result, the Ns for the degrees of freedom tended to be slightly larger, although none was greater than 400. As with TLS2, there was no indication in the data that the adequacy of the fit changed with increasing or decreasing N.

Results of the Kolmogorov-Smirnov goodness of fit tests for Bohj's TB1 were somewhat more complex than for the other statistics. Although the range of probability values was

TABLE 12

Kolmogorov-Smirnov p-values for the goodness of fit of TB1, TLS2, and SE : sample set 888

	TB1				TLS2			SE		
rho	df	N	p-value q=.1	p-value q=.9	df	N	p-value	df	N	p-value
.5	10	124	.06	.00	16	70	1.0	19	69	.25
	11	145	.04	.02	17	127	.31	20	96	.92
	12	120	.34	.02	18	148	.63	21	153	.95
	13	127	.01	.06	19	183	.65	22	233	.99
	14	99	.09	.80	20	213	1.0	23	338	.52
	15	103	.08	.01	21	129	.81			
.7	12	86	.05	.10	16	99	.99	18	80	.96
	13	99	.50	.08	17	132	.59	19	119	1.0
	14	132	.19	.05	18	165	1.0	20	140	1.0
	15	132	.09	.12	19	182	1.0	21	153	.98
	16	200	.00	.52	20	166	.95	22	204	.97
	17	235	.00	.11	21	84	.25	23	215	.22
.9	15	120	.66	1.0	13	102	.99	15	158	.53
	16	309	.60	.78	14	131	.19	16	228	.30
	17	542	.18	.04	15	223	1.0	17	225	.98
					16	189	1.0	18	179	1.0
					17	124	.82	19	84	1.0
					18	68	1.0			

NOTE: Only the six degrees of freedom with the largest N or
with N > 60 are reported here.
N = Number of samples.

TABLE 13

Kolmogorov-Smirnov p-values for the goodness of fit
of TB1, TLS2, and SE : sample set 8816

	TB1				TLS2			SE		
rho	df	N	p-value q=.1	 q=.9	df	N	p-value	df	N	p-value
.5	10	127	.01	.00	23	99	.59	27	69	1.0
	11	126	.00	.01	24	82	.85	28	82	.99
	12	119	.05	.00	25	93	1.0	29	137	1.0
	13	109	.01	.21	26	76	.96	30	246	.95
	14	80	.04	.06	27	88	.27	31	353	1.0
	15	79	.11	.28	28	98	.31			
.7	12	75	.28	.20	23	75	.96	26	96	1.0
	13	72	.36	.01	24	88	1.0	27	116	.32
	14	76	.55	.33	25	86	.46	28	138	1.0
	15	77	.83	.07	26	92	.99	29	155	1.0
	16	85	.08	.08	27	114	.65	30	185	.94
	17	68	.71	.43	28	103	.58	31	205	.72
.9	23	119	.17	.14	21	79	1.0	23	144	.44
	24	236	.92	.28	23	81	.73	24	264	1.0
	25	444	.35	1.0	24	90	.60	25	250	.96
					25	88	1.0	26	159	1.0
					26	69	1.0	27	94	1.0
					27	69	1.0			

NOTE: Only the six degrees of freedom with the largest N or
with N > 60 are reported here.
N = Number of samples.

TABLE 14

Kolmogorov-Smirnov p-values for the goodness of fit of TB1, TLS2, and SE : sample set 81616

	TB1				TLS2			SE		
rho	df	N	p-value		df	N	p-value	df	N	p-value
			q=.1	q=.9						
.5	10	124	.05	.00	32	93	1.0	35	72	.83
	11	115	.01	.00	33	103	1.0	36	111	1.0
	12	114	.02	.01	34	144	.64	37	135	.92
	13	96	.20	.42	35	165	.27	38	229	1.0
	14	74	.18	.11	36	205	1.0	39	340	1.0
	15	73	.24	.17	37	101	.10			
.7	13	68	.46	.07	32	94	1.0	34	102	1.0
	14	78	.35	.03	33	139	.88	35	119	1.0
	15	57	.40	.05	34	148	.69	36	144	.97
	16	61	.87	.45	35	157	.81	37	161	.86
	17	67	.04	.78	36	170	.97	38	175	.99
	18	57	.61	.33	37	92	.96	39	191	.63
.9	29	65	.72	.45	29	81	.82	31	128	.22
	30	71	.06	1.0	30	108	.98	32	298	.99
	31	110	.10	.98	31	202	1.0	33	226	.48
	32	168	.67	.95	32	230	1.0	34	152	.90
	33	238	.19	.69	33	134	.94	35	115	.82
					34	71	.61			

NOTE: Only the six degrees of freedom with the largest N or with N > 60 are reported here.
N = Number of samples.

TABLE 15

Kolmogorov-Smirnov p-values for the goodness of fit of TB1, TLS2, and SE : sample set 1688

	TB1				TLS2			SE		
rho	df	N	p-value q=.1	p-value q=.9	df	N	p-value	df	N	p-value
.5	22	92	.10	.46	24	91	.99	27	85	1.0
	23	139	.11	.01	25	105	.31	28	89	.87
	24	270	.00	.00	26	130	.94	29	136	.72
	25	416	.00	.00	27	180	.98	30	226	1.0
					28	201	.81	31	343	1.0
					29	105	.62			
.7	20	78	.06	.71	23	90	.68	26	80	.95
	21	114	.04	.12	24	92	1.0	27	86	.16
	22	122	.10	.19	25	122	.27	28	103	.73
	23	168	.24	.24	26	106	.65	29	128	.98
	24	182	.13	.11	27	110	.94	30	145	.00
	25	240	.27	.02	28	120	.04	31	208	.99
.9	16	138	.11	.02	15	78	.82	18	138	.61
	17	240	.01	.33	16	97	.82	19	153	.95
	18	199	.54	.62	17	154	.66	20	144	1.0
	19	137	1.0	.67	18	138	.71	21	98	.51
	20	104	1.0	.21	19	118	1.0	22	98	.88
	21	67	.52	.91	20	89	.07	23	75	.97

NOTE: Only the six degrees of freedom with the largest N or with N > 60 are reported here.
N = Number of samples.

TABLE 16

Kolmogorov-Smirnov p-values for the goodness of fit of TB1, TLS2, and SE : sample set 16816

	TB1				TLS2			SE		
rho	df	N	p-value		df	N	p-value	df	N	p-value
			q=.1	q=.9						
.5	28	64	.26	.03	31	68	.97	35	66	.59
	29	106	.44	.08	32	96	.71	36	89	.87
	30	104	.09	.05	33	110	.25	37	131	1.0
	31	142	.00	.05	34	95	.96	38	202	1.0
	32	168	.00	.10	35	122	.91	39	384	1.0
	33	231	.00	.11	36	125	1.0			
.7	30	64	.63	.98	31	76	1.0	34	89	.84
	31	119	.04	.60	32	78	1.0	35	105	.45
	32	274	.00	.01	33	94	1.0	36	110	.64
	33	450	.00	.01	34	94	.98	37	138	.08
					35	110	.50	38	138	.81
					36	131	.28	39	169	.97
.9	26	123	.01	.45	24	60	.67	26	165	.94
	27	147	.05	.41	26	79	1.0	27	140	.60
	28	146	.54	.99	27	79	.87	28	133	.98
	29	135	.40	.23	28	86	1.0	29	117	.98
	30	119	1.0	.09	29	80	1.0	30	88	1.0
	31	124	.84	.30				31	88	.36

NOTE: Only the six degrees of freedom with the largest N or with N > 60 are reported here.
N = Number of samples.

TABLE 17

Kolmogorov-Smirnov p-values for the goodness of fit of TB1, TLS2, and SE : sample set 161616

	TB1				TLS2			SE		
rho	df	N	p-value		df	N	p-value	df	N	p-value
			q=.1	q=.9						
.5	32	66	.17	.10	40	86	.46	44	82	.20
	34	73	.40	.14	41	95	1.0	45	142	.99
	35	85	.02	.08	42	140	.93	46	222	.80
	38	82	.05	.24	43	170	.25	47	370	.95
	39	69	.44	.16	44	201	.92			
	41	76	.22	.71	45	123	.99			
.7	37	89	.17	.18	39	77	.93	42	79	.35
	38	92	.74	.21	40	94	.26	43	100	.64
	39	138	.12	.09	41	118	1.0	44	115	.83
	40	196	.20	.33	42	152	.51	45	136	.57
	41	261	.00	.01	43	120	.90	46	180	1.0
					44	120	.63	47	161	.53
.9	36	102	.73	.21	32	87	.82	33	88	.99
	37	125	.59	.92	33	134	.99	34	136	1.0
	38	161	.27	.00	34	147	.83	35	177	.79
	39	183	.09	.25	35	146	.86	36	166	.76
	40	173	.32	.00	36	114	.99	37	130	.94
	41	152	.72	.47	37	69	.06	38	97	.22

NOTE: Only the six degrees of freedom with the largest N or with N > 60 are reported here.
N = Number of samples.

TABLE 18

Kolmogorov-Smirnov p-values for the goodness of fit of TB1, TLS2, and SE : sample set 2488

	TB1				TLS2			SE		
rho	df	N	p-value		df	N	p-value	df	N	p-value
			q=.1	q=.9						
.5	28	75	.20	.14	32	66	.98	36	85	.96
	29	101	.01	.03	33	110	.99	37	128	1.0
	30	120	.05	.37	34	127	.23	38	218	.51
	31	123	.06	.21	35	158	.96	39	388	.48
	32	133	.10	.06	36	217	.30			
	33	173	.28	.02	37	124	.79			
.7	21	82	.05	.34	31	73	.74	34	78	.43
	22	72	.13	.20	32	73	.23	35	66	1.0
	23	106	.04	.20	33	88	1.0	36	82	.99
	24	83	.03	.26	34	99	1.0	37	102	.97
	25	86	.39	.10	35	94	1.0	38	147	.84
	26	82	.41	.10	36	102	.75	39	197	.72
.9	16	156	.00	.88	17	73	.73	20	76	.42
	17	238	.08	.11	18	97	.36	21	118	.93
	18	190	.22	.02	19	105	.96	22	87	.21
	19	150	.80	.18	20	106	.65	23	81	.76
	20	82	.34	.65	21	87	.92	24	82	.92
	21	69	.36	.16	22	72	.77	25	83	.70

NOTE: Only the six degrees of freedom with the largest N or with N > 60 are reported here.
N = Number of samples.

TABLE 19

Kolmogorov-Smirnov p-values for the goodness of fit of TB1, TLS2, and SE : sample set 24816

	TB1				TLS2			SE		
rho	df	N	p-value q=.1	p-value q=.9	df	N	p-value	df	N	p-value
.5	38	88	.73	.09	40	83	.13	43	61	.77
	39	108	.05	.10	41	102	.92	44	73	.51
	40	255	.01	.06	42	121	.12	45	121	.99
	41	414	.00	.00	43	149	1.0	46	240	1.0
					44	147	.85	47	361	1.0
					45	74	.97			
.7	36	89	.20	.31	40	75	.19	41	69	.23
	37	90	.01	.61	41	92	.97	43	101	.15
	38	120	.10	.13	42	96	.71	44	93	.98
	39	125	.04	.11	43	108	.41	45	103	.99
	40	161	.86	.12	44	122	.90	46	140	.71
	41	197	.29	.05	45	67	.91	47	175	.71
.9	26	76	.21	.69	27	69	1.0	28	87	.38
	27	138	.08	.52	28	51	1.0	29	105	.26
	28	163	.50	.12	29	67	.58	30	116	.78
	29	140	.41	.49	30	79	1.0	31	119	.53
	30	126	.99	.68	31	76	.62	32	90	1.0
	31	93	.28	.03	32	68	.97	33	97	.28

NOTE: Only the six degrees of freedom with the largest N or with N > 60 are reported here.
N = Number of samples.

TABLE 20

Kolmogorov-Smirnov p-values for the goodness of fit of TB1, TLS2, and SE : sample set 241616

	TB1				TLS2			SE		
rho	df	N	p-value		df	N	p-value	df	N	p-value
			q=.1	q=.9						
.5	45	72	.62	.01	48	77	.99	52	97	.31
	46	80	.32	.15	49	120	.47	53	125	1.0
	47	124	.07	.05	50	135	.99	54	231	.86
	48	219	.01	.03	51	161	.81	55	363	.91
	49	305	.00	.06	52	202	.41			
					53	103	1.0			
.7	46	93	.17	.25	47	83	.18	50	75	.27
	47	130	.43	.18	48	88	1.0	51	83	.99
	48	234	.02	.02	49	97	.10	52	98	.74
	49	378	.08	.00	50	94	.99	53	116	.99
					51	103	.96	54	128	.86
					52	97	.96	55	138	.50
.9	37	114	.28	.85	34	78	1.0	36	106	.40
	38	116	.10	.87	35	123	1.0	37	115	1.0
	39	130	.19	.45	36	124	.24	38	133	.77
	40	121	.94	.09	37	110	.74	39	111	1.0
	41	114	.67	.27	38	111	.97	40	123	.40
	42	81	.09	.29	39	76	.05	41	90	.85

NOTE: Only the six degrees of freedom with the largest N or with N > 60 are reported here.
N = Number of samples.

similar, that is, from .00 to 1.0, many were less than .20 and very few were greater than .70. The values of the constant q, rho, and sample size affected the adequacy of the fit of TB1 to the appropriate t distribution.

Values of q=.1 and .9 consistently resulted in higher probability values than did values of q=.3, .5, or .7. The use of q=.1 produced a better fit for sample sets 8816, 81616, and 241616 when rho was equal to .7. The use of q=.9 produced a better fit for sample sizes 81616 and 16816 when rho was equal to .9. In all other cases however it was difficult to determine which of two values of q gave the best fit. Out of the six degrees of freedom with the largest N, three would have high probability values when q was equal to .1, while the remaining three would have high probability values with q equal to .9. This 'interaction' between degree of freedom and goodness of fit, for the most part, appeared to be arbitrary. In a few cases there was evidence of a pattern. In sample sets 16816, 161616, and 241616, when rho was equal to .5, a value of q=.1 produced the best fit for the three degrees of freedom with the small Ns, while a value of q equal to .9 produced the best fit for the three degrees of freedom with the large Ns. In sample sets 1688 and 24816, when rho was equal to .7, a value of q equal to .9 produced the best fit for the degrees of freedom with small Ns, while a value of q equal to .1 produced the best fit for the degrees of freedom with large Ns.

The goodness of fit of TB1 to its t distribution improved as the value of rho increased. With rho equal to .5 probability values ranged from .00 to .73 with q equal to .1, and from .00 to .80 with q equal to .9. The majority were less than .20 and many were less than .05. With rho equal to .9, probability values ranged from .00 to 1.0 for q equal to .1 and .9, with many values greater than .70. Sample sets 8816, 81616, and 1688, with rho equal to .9 produced probability values well within the range of the paired t. The remaining sample sets, on the whole, produced slightly lower probability values.

In some cases, the range of degrees of freedom for TB1 was fairly small, and Ns as large as 542 occurred. Whereas for TLS2 and SE the probability values did not appear to be related to the number of cases for a degree of freedom, with TB1 the degrees of freedom with the larger N's tended to have smaller probability values. Out of the twenty-two instances where the N for a degree of freedom was greater than 200, 11 probability values were less than .05 with q equal to .9, and 14 were less than .05 with q equal to .1. Approximately forty percent of the time the probability values decreased as N increased, within the degrees of freedom for a given sample set.

In summary, the goodness of fit probability values for TLS1 and ZE were judged to be acceptable: they were generally within the range of the paired t probability values, and none was less than .20. With only a few exceptions, the goodness of fit probability values for ZLS, TLS2, and SE were similarly judged to be acceptable. Those of TLS2 and SE tended to be slightly lower, overall, than those of the paired t, with a few falling below .20. Probability values for the fit of ZLS were judged to be unacceptable when the number of paired observations was small and the number of unpaired observations was large. For TB1, values of the constant q equal to .1 and .9 gave the best goodness of fit results. In most cases it was difficult to determine which of these two values was better, however, due to the large variation in probability values for the various degrees of freedom in a given sample set. For most conditions, the goodness of fit probability values for TB1 were very poor. Often they tended to decrease as the N increased within a given sample set. The probability values were only within the range of the paired t when n was small and rho was large.

In general, these goodness of fit results concur with those of Lin and Stivers who claimed that their statistics ZLS, TLS1, and TLS2 had acceptable fit (except for ZLS when n was small), as well as with those of Ekbohm, who claimed that the goodness of fit of SE and ZE to their respective distributions was accurate. Bohj did not examine the fit of his statistics.

SIZE

Tables 21 and 22 give the size (that is, the empirical alpha level) of all the statistics examined in this study, for the nominal alpha levels of .05 and .01. The figures in these Tables refer to the actual number of samples, out of the 1000 random samples generated for each combination of rho and sample size, in which the probability value of a statistic fell below .05 or .01. The size of a statistic was considered acceptable if it was within two standard deviations of the nominal alpha level (see page 18). In actual numbers (that is size X 1000) this was between 36 and 64 for the .05 nominal alpha level, and between 4 and 16 for the .01 nominal alpha level. Comparable figures for three standard deviations were between 29 and 71, and between 1 and 19.

The size figures for the paired t were within two standard deviations of the nominal alpha levels for all combinations of rho and sample size. The size figures for TLS1, TLS2, and SE were also within two standard deviations of the nominal alpha levels for all conditions of rho and delta. Size figures for ZE were on the whole slightly larger than those of the paired t, TLS1, TLS2, and SE under the conditions of high rho and small sample size. In three instances the size figures were beyond two standard deviations, although they were still within three standard deviations of the nominal alpha levels. For the nominal alpha level of .05, the size figures for ZLS were generally larger than those of all the other statistics for most combinations of rho and sample size. The size of ZLS exceeded three standard deviations of the .05 nominal alpha

TABLE 21

Size (Empirical alpha levels X1000) of all statistics with alpha=.05 and delta=0.0

Sample set	rho	t	ZLS	TLS1	TLS2	ZE	SE	TB1-.1	TB1-.9
888	.5	39	55	59	44	47	50	11**	26**
	.7	44	68*	45	45	56	47	21**	33*
	.9	58	83**	60	54	63	55	34*	47
8816	.5	45	57	49	44	37	41	7**	22**
	.7	46	78**	53	45	55	51	14**	28**
	.9	61	79**	55	51	70*	57	34*	59
81616	.5	47	59	54	47	50	48	6**	21**
	.7	52	77**	51	49	52	49	9**	32*
	.9	45	74**	50	53	59	55	31*	53
1688	.5	55	55	46	39	50	42	9**	8**
	.7	50	55	43	38	46	39	22**	18**
	.9	50	64	46	44	63	46	40	38
16816	.5	46	47	55	51	47	47	6**	12**
	.7	61	58	46	42	56	43	9**	24**
	.9	51	62	54	49	60	45	30*	35*
161616	.5	52	54	57	53	56	55	8**	14**
	.7	43	49	48	45	44	45	11**	13**
	.9	41	61	50	47	50	47	22**	28**
2488	.5	60	63	62	56	61	56	20**	12**
	.7	44	55	52	51	57	52	28**	12**
	.9	49	60	43	49	60	50	53	28**
24816	.5	51	59	60	56	55	54	3**	9**
	.7	42	46	57	53	45	50	18**	8**
	.9	50	53	54	59	55	58	34*	29*
241616	.5	64	61	53	51	56	51	9**	10**
	.7	47	56	42	43	53	45	10**	13**
	.9	50	52	44	45	54	45	34*	27**

* size departs from the nominal alpha level by 2 standard deviatons

** size departs from the nominal alpha level by 3 standard deviations

TABLE 22

Size (Empirical alpha levels X1000) of all statistics with alpha=.01 and delta=0.0

Sample set	rho	t	ZLS	TLS1	TLS2	ZE	SE	TB1-.1	TB1-.9
888	.5	9	14	8	4	10	4	0**	5
	.7	8	8	11	7	12	10	2*	12
	.9	7	17*	10	11	13	12	9	12
8816	.5	11	13	9	7	12	9	1*	10
	.7	4	13	10	10	15	8	4	9
	.9	8	22**	9	10	15	9	9	23**
81616	.5	7	9	12	9	9	9	1*	8
	.7	9	15	9	8	12	8	0**	8
	.9	7	14	13	12	12	12	7	17*
1688	.5	13	9	9	7	8	7	2*	1*
	.7	10	11	8	10	13	10	6	5
	.9	10	15	13	10	17*	10	11	10
16816	.5	11	13	10	10	13	11	1*	2*
	.7	8	12	8	9	10	11	1*	1*
	.9	10	16	14	15	18*	14	5	6
161616	.5	10	8	8	9	10	9	2*	2*
	.7	6	4	8	9	4	9	0**	1*
	.9	9	9	10	10	6	10	6	7
2488	.5	15	14	16	16	13	16	8	4
	.7	11	12	10	14	12	14	7	2*
	.9	15	14	14	13	12	13	18*	6
24816	.5	8	7	12	11	9	12	0**	1*
	.7	8	8	12	11	6	12	4	2*
	.9	11	14	11	11	12	12	9	3*
241616	.5	10	11	15	15	9	13	0**	0**
	.7	13	10	13	12	10	12	2*	1*
	.9	10	8	8	8	8	9	2*	3*

* size departs from the nominal alpha level by 2 standard deviations

** size departs from the nominal alpha level by 3 standard deviations for high values of rho (.7 and .9) in combination with small numbers of paired observations (n=8). The same trend occurred in size figures for the nominal alpha level of .01, but to a lesser extent. Only two size figures actually exceeded two standard deviations of the nominal alpha level.

As was the case with goodness of fit, the results with regard to size for TB1 were more complex than for the other statistics examined in this study. All values of the constant q gave fairly small size figures. Values of q equal to .3, .5, and .7 consistently gave much smaller size figures than did values of q equal to .1 and .9. Only the latter are therefore reported in Tables 21 and 22. Of the two, a value of q equal to .1 gave the best size results when n, the number of paired observations, was equal to or greater than the sum of j plus k, the number of unpaired observations, and/or when both sample size and rho were large. Otherwise a value of q equal to .9 gave the best size results.

The size figures for TB1 were on the whole much smaller than those of the other statistics for the .05 nominal alpha level when rho was equal to .5 and .7. All but one exceeded three standard deviations. With rho equal to .9, and for small values of n, at least one value of q gave size results more within the range of those of the other statistics. In all but one instance, sample set 161616, at least one value of q gave size results within three standard deviations of the nominal .05 level. Size figures for TB1 were somewhat better for the .01 nominal alpha level, although in a few instances when rho was equal to .9, the size exceeded three standard deviations. With rho equal to .5 in sample sets 888(q=.1), 24816(q=.1) and 241616(q=.1 and .9), and with rho equal to .7 in sample set 161616(q=.1), the size of TB1 was too small. With rho equal to .9 in sample set 8816(q=.9) size was too large. For the smaller sample sizes, the size figures tended to be within the range of the other statistics. For larger sample sizes, size figures tended to be smaller than those of the other statistics. Again, in only one instance (rho=.5 in sample set 241616), did both values of q fail to give a size value within three standard deviations of the .01 nominal alpha level. The overall range of size figures for TB1 was greater than that of the other statistics, since, despite a large number of extremely small size figures, there were also a few relatively large ones.

In summary, the size of TLS1, TLS2, and SE was judged to be good as all values were within two standard deviations of the nominal alpha level. The size of ZE was similarly judged to be good except for those samples sets where the number of paired observations was small and rho was high. The size of ZLS was judged to be fair only, since size figures on the whole tended to be larger than those of the paired t, TLS1, TLS2 and SE. For TB1, values of q equal to .9 and .1 gave the best size results. Size figures for TB1 showed a greater range than those of the other statistics: most were too small (when rho was equal to .5 and .7) while a few were too large (when rho was equal to .9). For the .05 nominal alpha level the size of TB1 was poor for the most part. At alpha equal to .01, at least one value of q resulted in size figures within the fair to good range. This difference in results at the .05 and .01 alpha levels is consistent with the poor goodness of fit results for TB1.

As was the case with goodness of fit, these results agree with the results given in the literature, in that the size of ZLS, TLS1, TLS2, SE and ZE were good under the conditions examined in this study, with the exception that ZLS was too large when n was small. Bohj did not examine the size of TB1.

POWER

Power results for all statistics are given in Tables 23 to 26. The power of each of the new statistics was compared to the power of the paired t under varying conditions of rho, n, j, and k, for two conditions of delta(.4 and .8), and for the two nominal alpha levels of .05 and .01. Power was calculated as the actual number, out of the 1000 random samples generated for each condition, of probability values that fell below .05 and .01. Given a true mean difference of .4 or .8, the power was the actual number of times out of 1000 that the null hypothesis of delta=0 was correctly rejected.

With regard to sample size, the power of all the new statistics increased as n increased and as j + k increased

TABLE 23

Power (X1000) of all statistics
with alpha=.05 and delta=0.4

Sample set	rho	t	ZLS	TLS1	TLS2	ZE	SE	TB1-.9
888	.5	480	684	695	683	686	254	295
	.7	719	837	767	743	833	271	536
	.9	995	994	847	819	987	304	976
8816	.5	508	769	765	750	743	248	324
	.7	705	886	818	802	859	301	553
	.9	989	995	835	816	983	285	973
81616	.5	518	825	858	859	820	335	329
	.7	711	936	886	878	919	373	575
	.9	993	998	894	890	994	370	983
1688	.5	838	916	911	910	913	378	509
	.7	977	988	960	957	984	452	854
	.9	1000	1000	994	995	1000	524	1000
16816	.5	853	937	930	928	934	418	598
	.7	982	992	960	960	989	447	878
	.9	1000	1000	990	988	1000	537	1000
161616	.5	848	961	957	958	964	445	569
	.7	972	994	975	972	994	520	885
	.9	1000	1000	993	987	1000	544	1000
2488	.5	964	983	984	983	984	485	731
	.7	997	999	998	998	999	627	969
	.9	1000	1000	1000	1000	1000	762	1000
24816	.5	962	989	986	986	989	549	768
	.7	988	1000	995	999	1000	620	985
	.9	1000	1000	1000	1000	1000	733	1000
241616	.5	952	990	983	985	991	540	773
	.7	997	999	995	993	999	669	984
	.9	1000	1000	999	999	1000	719	1000

TABLE 24

Power (X1000) of all statistics
with alpha=.01 and delta=0.4

Sample set	rho	t	ZLS	TLS1	TLS2	ZE	SE	TB1-.9
888	.5	217	389	476	449	379	100	131
	.7	378	562	516	479	506	102	310
	.9	897	959	639	572	907	113	867
8816	.5	226	439	517	491	409	97	156
	.7	374	634	602	581	577	131	315
	.9	885	953	606	568	900	110	880
81616	.5	220	529	646	630	504	145	160
	.7	339	695	704	689	667	161	318
	.9	892	972	718	695	922	157	899
1688	.5	598	735	733	709	730	158	243
	.7	862	922	854	835	910	210	593
	.9	1000	1000	961	933	1000	273	996
16816	.5	638	790	780	773	772	214	243
	.7	880	955	858	845	937	228	645
	.9	999	1000	953	930	1000	281	998
161616	.5	607	847	852	846	841	229	286
	.7	852	960	922	908	953	261	673
	.9	1000	1000	945	946	1000	282	1000
2488	.5	860	927	909	900	920	257	399
	.7	984	996	977	972	992	352	822
	.9	1000	1000	1000	995	1000	527	1000
24816	.5	856	938	927	919	942	295	475
	.7	991	995	979	976	994	320	869
	.9	1000	1000	999	998	1000	494	1000
241616	.5	854	947	939	932	949	277	475
	.7	988	996	973	968	991	398	876
	.9	1000	1000	999	996	1000	470	1000

TABLE 25

Power (X1000) of all statistics
with alpha=.05 and delta=0.8

Sample set	rho	t	ZLS	TLS1	TLS2	ZE	SE	TB1-.9
888	.5	969	1000	999	1000	999	696	803
	.7	996	1000	1000	1000	1000	776	968
	.9	1000	1000	1000	1000	1000	804	1000
8816	.5	965	1000	1000	1000	1000	752	826
	.7	997	1000	1000	1000	1000	806	970
	.9	1000	1000	1000	1000	1000	857	1000
81616	.5	973	1000	1000	1000	1000	831	832
	.7	999	1000	1000	1000	1000	874	974
	.9	1000	1000	1000	1000	1000	866	1000
1688	.5	1000	1000	1000	1000	1000	905	991
	.7	1000	1000	1000	1000	1000	967	1000
	.9	1000	1000	1000	1000	1000	984	1000
16816	.5	1000	1000	1000	1000	1000	922	994
	.7	1000	1000	1000	1000	1000	966	1000
	.9	1000	1000	1000	1000	1000	982	1000
161616	.5	1000	1000	1000	1000	1000	958	997
	.7	1000	1000	1000	1000	1000	981	1000
	.9	1000	1000	1000	1000	1000	989	1000
2488	.5	1000	1000	1000	1000	1000	980	1000
	.7	1000	1000	1000	1000	1000	993	1000
	.9	1000	1000	1000	1000	1000	1000	1000
24816	.5	1000	1000	1000	1000	1000	979	1000
	.7	1000	1000	1000	1000	1000	992	1000
	.9	1000	1000	1000	1000	1000	999	1000
241616	.5	1000	1000	1000	1000	1000	989	1000
	.7	1000	1000	1000	1000	1000	996	1000
	.9	1000	1000	1000	1000	1000	1000	1000

TABLE 26

Power (X1000) of all statistics
with alpha=.01 and delta=0.8

Sample set	rho	t	ZLS	TLS1	TLS2	ZE	SE	TB1-.9
888	.5	793	964	985	983	963	417	532
	.7	960	999	997	995	998	499	825
	.9	999	1000	1000	1000	996	558	998
8816	.5	795	985	1000	994	985	499	570
	.7	945	997	1000	998	996	571	848
	.9	1000	1000	1000	1000	1000	662	1000
81616	.5	801	994	1000	1000	997	624	581
	.7	956	1000	1000	1000	999	673	856
	.9	1000	1000	1000	1000	999	664	1000
1688	.5	997	1000	999	1000	1000	729	911
	.7	1000	1000	1000	1000	1000	846	999
	.9	1000	1000	1000	1000	1000	929	1000
16816	.5	998	1000	1000	1000	1000	770	921
	.7	1000	1000	1000	1000	1000	869	1000
	.9	1000	1000	1000	1000	1000	910	1000
161616	.5	999	1000	1000	1000	1000	834	941
	.7	1000	1000	1000	1000	1000	894	999
	.9	1000	1000	1000	1000	1000	920	1000
2488	.5	1000	1000	1000	1000	1000	902	993
	.7	1000	1000	1000	1000	1000	960	1000
	.9	1000	1000	1000	1000	1000	997	1000
24816	.5	1000	1000	1000	1000	1000	921	994
	.7	1000	1000	1000	1000	1000	963	1000
	.9	1000	1000	1000	1000	1000	993	1000
241616	.5	1000	1000	1000	1000	1000	935	993
	.7	1000	1000	1000	1000	1000	969	1000
	.9	1000	1000	1000	1000	1000	990	1000

within a value of n. The increase in power across the three values of n was greater than the increase across the three values of j + k. For the paired t, power also increased as n increased. Within a given value of n, however, there was no increase in power as j + k increased, as would be expected. Although all the new statistics were more sensitive to an increase in the number of paired observations than to an increase in the number of unpaired observations, the presence of an increase in power with the increase in j + k does indicate that, to a greater or lesser degree, all the new statistics were making use of the unpaired data to increase power.

The power of all statistics increased as the value of rho increased, indicating once more that they all are particularly sensitive to the paired data. Of all the new statistics, SE seemed to be the least sensitive to increases in rho (that is, the increase in power over the three values of rho was smaller than that of the other statistics), while TB1 seemed to be the most sensitive to an increase in rho.

As would be expected, the power of all statistics was greater for delta equal to .8 than for delta equal to .4. Unfortunately, for values of delta equal to .8, with alpha equal to .05, as well as for the larger sample sizes with alpha equal to .01, there was a substantial ceiling effect for all statistics except SE and TB1. Power results for delta equal to .4 and alpha equal to .01 (Table 24) provide the clearest comparison of the power of the statistics.

Power figures for the paired t ranged from a low of 217 for rho equal to .5 in sample set 888 (delta=.4 and alpha=.01), to a high of 1000 for a number of conditions with large sample size and/or high rho. The power of the paired t decreased as rho, alpha, delta, and n decreased until with delta equal to .4, power figures were poor (that is, less than .75) for 1) rho equal to .5 and .7 and n equal to 8 with alpha equal to .05 and .01, and 2) rho equal to .5 and n equal to 16 for alpha equal to .01.

In most cases, power figures for ZLS and ZE were about the same. Only in Table 24 with delta equal to .4 and alpha equal to .01 are the power figures for ZLS consistently larger than those of ZE for all values of rho and the smaller sample sizes. Both of these statistics gave good (that is, greater than .75) power figures for the widest variety of conditions. Power was poor for these two statistics only under the conditions of alpha equal to .01 and delta equal to .4 with n equal to 8 and rho equal to .5 and .7. Of the new statistics, power figures for ZLS and ZE, the two based on maximum likelihood estimates of the true mean difference, were higher than or approximately the same as those of the paired t for all conditions. Both ZLS and ZE were substantially more powerful than the paired t for low values of rho combined with small sample sizes. As rho increased and sample size increased, the gain in power for these two statistics over that of the paired t decreased. With rho equal to .9 and sample sizes greater than 16,8,8 (for delta=.4 and alpha=.05 and .01), there were no differences in power among ZLS, ZE, and the paired t. With delta equal to .8 and alpha equal to .01, all samples with rho equal to .9 showed similar power figures for ZLS, ZE, and the paired t; while with delta equal to .8 and alpha equal to .05, all samples with rho equal to .7 and .9 gave similar figures for ZLS, ZE and the paired t. Thus, although ZLS and ZE showed the best power results overall, they were only more powerful than the paired t for 1) low values of rho, 2) small sample size, and 3) small values of delta.

On the whole, power figures for TLS1 and TLS2 were lower than those of ZLS and ZE. In most cases power figures for TLS1 tended to be slightly higher than those of TLS2, although both statistics seemed to perform the same in relation to all the other statistics. Power was poor for these two statistics (less than .75) for all values of rho when alpha was equal to .01, delta was equal to .4, and n was equal to 8. TLS1 and TLS2 did have higher power values than all the other statistics however for all values of delta and alpha when rho was equal to .5 and n was equal to 8. With rho equal to .5 and n equal to 16 and 24, the power of these two statistics was similar to that of ZLS, ZE and the paired t. As rho increased and as sample size increased, TLS1 and TLS2 became less powerful than the paired t. Thus although the power figures for TLS1 and TLS2 were better than or the same as those of the paired t under a number of conditions, these two statistics were only more powerful than all the other statistics when rho was small and the number of paired observations was also small.

The power of TB1 was good (greater than .75) under even fewer combinations of conditons: specifically 1) when rho was equal to .9, regardless of the values of the other parameters, and 2) for all values of rho, n, and alpha with delta equal to .8(except n=8, rho=.5 and alpha=.01). As n and rho decreased, power decreased until it was poor for the lower values of rho and n with delta equal to .4 and alpha equal to .05 and .01. The power figures for TB1 were generally lower than those of ZLS, ZE, TLS1, TLS2, and the paired t. With delta equal to .8 the power of TB1 increased with increasing n and increasing rho until, for rho equal to .7 and n equal to 16 and 24 (alpha=.05 and .01), as well as for rho equal to .9 and all values of n, the power figures for TB1 were comparable to those of ZLS, ZE, TLS1, TLS2, and the paired t. With delta equal to .4, power figures for TB1 were similar to these other statistics for values of rho equal to .9 and n equal to 16 and 24(alpha =.05), and for rho equal to .9 with n equal to 24 (alpha=.01). There was no combination of conditions under which TB1 was more powerful than all the other statistics.

Power figures were consistantly the lowest for SE. Only with delta equal to .8 and alpha equal to .05 was the power good (greater than .75) for all values of n and rho. For a given combination of delta and alpha, the power figures became poorer as n and rho decreased. They were less than

.75 for all values of rho, sample size, and alpha when delta was equal to .8. Compared to the other statistics, SE was considerably less powerful under the conditions of small n and large rho. Within a given set of conditions, the increase in power due to an increase in rho, was relatively small for SE. As sample size increased, and as delta increased, power figures for SE increased substantially. With delta equal to .8, rho equal to .7 and .9, and n equal to 24 (alpha=.05 and .01), power figures for SE were comparable to those of the other statistics. There were no conditions under which SE was more powerful than any of the other statistics.

In summary, none of the statistics examined in this study, including the paired t, had especially good power for conditions of low rho combined with small numbers of paired observations, especially with alpha equal to .01 and delta equal to .4. They all showed increasingly better power as delta, rho, and sample size increased. The gains in power over the paired t tended to be greatest for low n. These gains decreased as n, rho, and delta increased. With rho equal to .9 and n equal to 16 and 24, power gains of any statistic over the paired t were small. All the statistics were more sensitive to an increase in the number of paired observations than to an increase in the number of unpaired observations, although an increase in the number of unpaired observations did increase power noticeably. Of all the statistics, SE was the most and TB1 the least sensitive to an increase in the value of rho.

ZLS and ZE were similar in power. They were as powerful as or more powerful than the paired t under all of the conditions examined in this study. They tended to be the most powerful statistics for the larger sample sizes. Power figures for TLS1 and TLS2 were similar to each other, with a slight advantage for TLS1. The power of these two statistics was slightly lower than that of ZE and ZLS under the majority of conditions. Overall their power was as good as or better than the paired t. TLS1 and TLS2 were the most powerful statistics for small numbers of paired observations combined with low values of rho. Power values for TB1 were best for values of the constant q equal to .9. The power of TB1 was only judged to be good when rho was high. Except for the ceiling effect on the other statistics when delta was equal to .8, power values for TB1 were lower than those of the paired t and the other statistics (except SE), for low values of rho and sample size. The power of SE was consistently the lowest of all the satistics. It was judged to be good only when the true mean difference was large and alpha was equal to .05. There was no set of conditions under which either TB1 or SE was more powerful than the paired t or than ZE, ZLS, TLS1 or TLS2.

These results are substantiated (except for TB1, and to a lesser extent, SE) by the literature. Lin and Stivers concluded that TLS1 or TLS2 were the best of their statistics for conditions of low rho, and ZLS was best for conditions of high rho and large n. Ekbohm found ZLS and ZE to be the most powerful of the statistics he examined. He also found TLS1 to be more powerful than TLS2 for conditions of equal variances. SE was never the most powerful statistic, although Ekbohm's power figures for SE were higher than those found in the present study. Ekbohm recommends ZE and ZLS for high values of rho (greater than .5) and TLS1, TLS2 and SE for low values of rho. Bohj did not explicitly examine the power of his statistics. By comparing the expected length squared of the confidence intervals of TB1 and the paired t, Bohj was implicitly examining the power of his statistic. Results for his statistic TB1 improved as rho decreased and as j and k increased, whereas in this study, power figures for TB1 increased as rho increased and as n increased - with results never being better than the paired t.

V. DISSCUSION, CONCLUSIONS, AND IMPLICATIONS

A. DISSCUSION

The main purpose of this study was to examine several new statistics which have been designed to give better results than the paired t in the pre-test post-test design with missing observations. These statistics were evaluated and compared with each other and with the paired t in terms of 1) the goodness of fit to the appropriate t distribution, 2) size (empirical alpha level with delta equal to 0.0), and 3) power (the proportion of null hypotheses correctly rejected when a true difference exists). In earlier studies by Lin and Stivers, and by Ekbohm, five of these statistics, ZLS, ZE, TLS1, TLS2, and SE, showed some promise under specific sets of conditions. These studies reported data regarding the size and power of these statistics, with only passing reference to the goodness of fit of each statistic to its t distribution. In his study, Bohj did not examine the size or goodness of fit of his statistics. He reported on the expected size of the confidence limits of his statistics in comparison to the paired t (and thus implicitly examined power). It was not possible however to directly compare the performance of Bohj's statistics to the other statistics. The specific direction of this study therefore, was to examine the performance of six of these new statistics, all calculated from the same data, using the same performance criteria (fit, size, and power), and under a set of conditions most relative to educational research.

B. CONCLUSIONS

EVALUATION OF THE SIX NEW STATISTICS

EKbohm's statistic ZE showed the best overall performance under the greatest variety of conditions. The goodness of fit of ZE to its t distribution was as good as the fit of the paired t to its t distribution. The size of ZE tended to be slightly larger than the paired t (particularly for high values of rho and low numbers of paired observations), but overall only three size figures were between two and three standard deviations of the nominal alpha level. The power of ZE was good for all conditions of rho and sample size when the true mean difference was large. With delta equal to .4 and alpha equal to .01 power was poor for low values of rho and small numbers of paired observations. Under these circumstances however, the power gains over the paired t were great. Although ZE was rarely the one best statistic, its power was as good as or better than the paired t under the widest variety of conditions.

The performance of ZLS was similar to that of ZE - the major difference being that with small numbers of paired observations either the size or fit of ZLS was poor for high values of rho and small numbers of unpaired observations when the true mean difference was small. With n greater than

8, ZLS was the most powerful statistic of those examined in the study.

The fit and size of TLS1 were good under all of the conditions examined in this study. When the true mean difference delta was large, the power of TLS1 was good for all combinations of sample size, rho, and alpha. With a small delta, power decreased as rho and sample size decreased. With alpha equal to .01 and small numbers of paired observations, the power of TLS1 was poor. All the statistics had low power under these circumstances, however, and TLS1 was more powerful than the others when n was equal to 8 and rho was equal to .5.

The fit and the size of TLS2 was also good under all the conditions examined in this study. The power of TLS2 closely paralleled that of TLS1. TLS2 was slightly less powerful than TLS1 except when the number of unpaired observations was large and rho was low. In this case, both ZE and ZLS were more powerful than TLS2, however.

The fit and size of SE were good under all conditions. The power of SE was generally quite poor and much lower than that of all the other statistics. Only with a large true mean difference and a large number of paired observations, did the power of SE begin to be acceptable and as good as the other statistics.

Bohj's statistic TB1 performed relatively poorly on all three criteria. The fit of TB1 was only consistently good for all of the six degrees of freedom with the highest N when rho was equal to .9 and n was equal to 8, although all sample sizes with rho equal to .9 had the majority of degrees of freedom with acceptable probability values for goodness of fit. With alpha equal to .05, the size of TB1 was fair to good for all sample sizes when rho was equal to .9 only. Size figures were better for alpha equal to .01, where most figures were fair to good. The power of TB1 was generally good when delta was equal to .8 (except for low rho combined with low n). With delta equal to .4, only rho equal to .9 showed good power. Although the power of TB1 was often as good as that of the other statistics (with high rho and high n), this was primarily because the other statistics had reached the ceiling value of 1000. Overall Bohj's statistic TB1 was totally unacceptable compared to the other statistics. In addition to poor performance on the fit, size, and power criteria singly, for a given combination of conditions it was usually difficult to select one value of the constant q which would maximize performance on all three criteria simultaneously.

COMPARISONS WITH THE PAIRED t

Table 27 presents information as to when and with which statistics, significant gains in power can be expected over the paired t. A difference of five points (that is, .005) was considered to be a significant gain in power. A statistic was not included for those sets of conditions where either its size or fit was considered unacceptable.

TABLE 27

Statistics with significant power gains over the paired t (in decreasing order)

Sample set	rho	alpha=.05,delta=.4	alpha=.01,delta=.4
888	.5	TLS1,ZE,ZLS,TLS2	TLS1,TLS2,ZLS,ZE
	.7	ZR,TLS1,TLS2	ZLS,TLS1,ZE,TLS2
	.9	---	ZE
8816	.5	TLS1,TLS2,ZE	TLS1,TLS2,ZE
	.7	ZE,TLS1,TLS2	ZLS,TLS1,TLS2,ZE
	.9	---	ZE
81616	.5	TLS2,TLS1,ZLS,ZE	TLS1,TLS2,ZLS,ZE
	.7	ZE,TLS1,TLS2	TLS1,TLS2,ZE
	.9	---	ZLS,ZE
1688	.5	ZLS,ZE,TLS1,TLS2	ZLS,TLS1,ZE,TLS2
	.7	ZLS,ZE	ZLS,ZE
	.9	---	---
16816	.5	ZLS,ZE,TLS1,TLS2	ZLS,TLS1,TLS2,ZE
	.7	ZLS,ZE	ZLS,ZE
	.9	---	---
161616	.5	ZE,ZLS,TLS2,TLS1	TLS1,ZLS,TLS2,ZE
	.7	ZLS&ZE	ZLS,ZE,TLS1,TLS2
	.9	---	---
2488	.5	TLS1&ZE,TLS2&ZLS	ZLS,ZE,TLS1,TLS2
	.7	---	ZLS,ZE
	.9	---	---
24816	.5	ZLS&ZE,TLS1&TLS2	ZE,ZLS,TLS1,TLS2
	.7	ZLS&ZE,TLS2,TLS1	ZE
	.9	---	---
241616	.5	ZLS,ZE,TLS1,TLS2	ZE,ZLS,TLS1,TLS2
	.7	---	ZLS
	.9	---	---

Although ZE was the one best statistic under the widest variety of conditions, Table 27 shows that the most significant gains in power over the paired t were obtained by using other statistics, according to the value of rho and sample size. Specifically, if the population correlation was low (rho=.5), TLS1 was the preferred statistic if the number of paired observations was small (less than 16), and ZLS was the preferred statistic if the number of paired observations was large. When the population correlation was .7, no one statistic was preferred over the paired t when the number of paired observations was small. Each of the statistics ZE, ZLS, and TLS1 were best in turn. When the number of paired observations was large, ZLS was the preferred statistic. When the population correlation was high (rho=.9) none of the statistics was preferred over the paired t, if the number of paired observations was large. If the number of paired observations was small (n=8), ZE showed slight gains in power over the paired t.

C. IMPLICATIONS

Of the six new statistics, only three, TLS1, ZLS, and ZE showed promise as statistics to be used in preference to the paired t under some of the conditions examined in this study. Since the performance of these statistics was basically acceptable with regard to fit and size, the major advantage of one statistic over another was due to larger gains in power over the paired t. When the number of paired observations was greater than eight, ZLS was the most powerful statistic - except when rho was equal to .9, in which case no statistic was much more powerful than the paired t. When the number of paired observations was small, TLS1 was the most powerful statistic when rho was equal to .5. As rho increased to .7, ZE and ZLS were more powerful than the paired t as often as was TLS1. For rho equal to .9 gains in power over the paired t were poor, with ZE and ZLS showing slight gains only with alpha equal to .01 and delta equal to .4. The performance of TLS2 was similar to that of TLS1, only slightly poorer in most cases. However, there are indications in the literature that TLS2 may be the preferred statistic under certain conditions of rho and sample size if the variances of X and Y are quite different. The performance of SE was particularly poor with regard to power and therefore showed no promise over the paired t. In his study, Bohj did not explicitly examine TB1 in terms of fit, size, or power (although the examination of the squared length of its confidence interval implies power). In this study, the performance of TB1 in comparison to the paired t and to the other statistics was disappointingly poor on all three criteria.

The above conclusions apply primarily to the condition of delta equal to .4. With a larger true mean difference (delta=.8) none of the statistics was more powerful than the paired t for any values of rho, unless the number of paired observations was small and rho was small to medium (.5 or

.7): in which case TLS1 generally offered the most gain in power over the paired t.

Cohen(1977) considers a true mean difference of .5 as a medium effect size, that is, a difference between groups which a person close to a situation could become aware of without formal measurement. Anything larger would be relatively obvious, while anything smaller would be more typically found in research conducted in new areas of inquiry, where the true difference may be small due to lack of control over extraneous variables. Relating this to the present findings, it becomes evident that the advantage of using the most powerful of the new statistics is primarily under such conditions, that is when the true mean difference is expected to be small. When the true mean difference is obvious to a trained observer, the gains in power over the paired t are slight unless sample size is very small.

Taking all of the above into consideration, one can conclude that, when using the pre-test post-test paradigm, there are a limited number of circumstances in which one of the newly developed statistics can offer large gains in power over the paired t. Attention must be given to the possible values of rho and delta, as well as to sample size. As rho and delta increase and interact, the usefulness of the new statistics decreases. It is only if the researcher feels that 1) the value of the population correlation is not too high (that is, less than .9) and that 2) the true mean difference is small and not obvious, that one of the new statistics could be applied to the data with the assurance that the results will be more powerful than using the paired t. If such is the case, the absolute gain in power will be greatest for small sample size.

VI. BIBLIOGRAPHY

Bohj,D.S.; Testing equality of means of correlated variates with missing observations on both responses, Biometrika, 1978, 65 (1), 225-8.

Cohen,J.; Statistical Power Analysis for the Behavioural Sciences, New York: Academic Press, 1977.

Ekbohm,G.; On comparing means in the paired case with incomplete data on both responses, Biometrika, 1976, 62 (2), 299-304.

Ekbohm,G.; Comparing means in the paired case with missing data on one response, Biometrika, 1976, 63 (1), 169-72.

Ghosh,B.K.; On the distribution of the difference of two t-variables, Journal of the American Statistical Association, 1975, 70 (350), 463-67.

Henkel,R.E.; Tests of Significance, Beverly Hills: Sage Publications , 1976.

International Mathematical and Statistical Libraries; Reference Manual , Houston, Texas: IMSL, Inc., 1980.

Kaiser,H., and Dickman,K.; Sample and Population Score Matrices from an Arbitrary Population Correlation Matrix, Psychometika, 1962, 27 (2), 179-182.

Kendall,M.G., and Stuart,A.; The Advanced Theory of Statistics, London: Charles Griffen and Co. Ltd., 1967.

Lee,A.F.S.; Size and power of tests for equality of means of two normal populations with unequal variances, Journal of the American Statistical Association, 1973, 68 (343), 933-41.

Lin,P.E.; Estimation procedures for differences of means with missing data, Journal of the American Statistical Association, 1971, 66 (335), 634-6.

Lin,P.E.; Procedures for testing the difference of means with incomplete data, Journal of the American Statistical Association, 1973, 68 (343), 699-703.

Lin,P.E., and Stivers,L.E.; On differences of means with incomplete data, Biometrika, 1974, 61 (2), 325-34.

Lin,P.E. and Stivers,L.E.; Testing for equality of means with incomplete data on one variable: a Monte Carlo Study, Journal of the American Statistical Association, 1975, 70 (349), 190-3.

Lord,F.M.; Estimation of parameters from incomplete data, Journal of the American Statistical Association, 1955, 55 (271), 870-6.

Mehta,J.S. and Gurland,J.; Some Properties and an application of a Statistic arising in Testing Correlation, The Annals of Mathematical Statistics, 1969a, 40 (5), 1736-45.

Mehta,J.S. and Gurland,J.; Testing equality of means in the presence of correlation, Biometrika, 1969a, 56 (1), 119-26.

Mehta,J.S. and Gurland,J.; A test for equality of means in the presence of correlation and missing values, Biometrika, 1973, 60 (1), 211-3.

Morrison,D.F.; Expectations and variances of maximum likelihood estimates of the multivariate normal distribution parameters with missing data, *Journal of the American Statistical Association*, 1971, *66* (335), 602-4.

Morrison,D.F.; A test for equality of means of correlated variates with missing data on one response, *Biometrika*, 1973, *60* (1), 101-5.

Naik,U.D.; On testing equality of means of correlated variables with incomplete data, *Biometrika*, 1975, *62* (3), 615-22.

Patil,V.H.; Approximations to the Behrens-Fisher distributions, *Biometrika*, 1965, *52* , 267-71.

Siegal,S; *Nonparametric Statistics for the Behavioural Sciences*, Tokyo: McGraw-Hill Kogakusha Ltd, 1956.

VII. APPENDIX I

The following program was run under double precision Fortran IV on the Amdahl 470 V/7 computer at the University of Alberta, using the Michigan Terminal System operating system. Subroutines GGUBFS and MDTD are from the IMSL Library. The approximate time and cost for running 1000 samples of 48 (sample set 161616) from a population of 3000 was .816 minutes for $1.45 at deferred priority. The program is limited to a total sample size of 100 and a population of 3000.

Given two populations previously generated (read from device #9) with specific means, variances, and correlation, this program is designed to select any number of samples of a desired size, and treat specific portions of these samples as missing observations. The statistics ZLS, TLS1, TLS2, ZE, SE, TB1, and the paired t are calculated for each sample. The power or size (depending on the value of delta) of each satistic, over the total number of samples, is also calculated. Output includes a summary of input parameters, Fisher's r transformation of the population correlation coefficient, the ranges of degrees of freedom for those statistics with variant degrees of freedom, the size or power of each statistic at the .05 and .01 alpha levels, and the final seed. In addition, for each sample, the means of the paired and unpaired data for X and Y variables, X and Y variances, and the actual value of each statistic, is written into a file (on device #10) for further analysis of the distributions of same.

```
      IMPLICIT REAL*8 (A-H,$,P-Z)
      DIMENSION X(100),Y(100),FMTD(20),XP(3000),YP(3000),
     *HT(5),VL(5),OP(5),NSP5(5),NSP1(5),OHT(5)
      READ(5,1)(VL(I),I=1,5),D,R,N,J,K,NT,NPOP,DSEED
    1 FORMAT(5F5.1,2F5.2,5I5,D16.9)
      IF(DSEED.EQ.0.0D0)READ(11,850)DSEED
  850 FORMAT(1X,D16.9)
      WRITE(6,1)(VL(I),I=1,5)
      WRITE(6,2)D,R,N,J,K,NT,NPOP,DSEED
    2 FORMAT(' ','DELTA=',F5.2,/,
     *'CORRELATION=',F5.2,/,
     *'NO. OF PAIRED OBSERVATIONS=',
     *I5,/,'NO. OF UNPAIRED OBSERVATIONS ON X=',I5,/,
     *'NO. OF UNPAIRED OBSERVATIONS ON Y=',I5,/,
     *'NO. OF SAMPLES TO BE GENERATED=',I5,/,
     *' SIZE OF POPULATION=',I6,/,'SEED=',D16.9)
      ZRHO=DLOG(DSQRT((1.0D0+R)/(1.0D0-R)))
      WRITE(6,801)ZRHO
  801 FORMAT(' Z TRANSFORMATION OF RHO=',F15.8)
      VNPOP=NPOP
      NPTPOW=0
      NUTPOW=0
      NUAPOW=0
      NPOW4=0
      NPOW2=0
      NPOW3=0
      NPZLS5=0
      NPZLS1=0
      NPLS25=0
      NPLS21=0
      NPLS45=0
      NPLS41=0
      NPZE5=0
      NPZE1=0
      NPSE5=0
      NPSE1=0
      DO 700 I=1,5
      NSP5(I)=0
      NSP1(I)=0
  700 CONTINUE
      NN=N+J+K
      N1=N+1
      N2=N+J+1
      N3=N+J
      VN=NN
      VNPT=N-1
      VNUPT=J+K-2
      ZN=N
      ZJ=J
      ZK=K
      DFTLS2=ZN+ZK+ZJ-4.0D0
      WRITE(6,208)ZN,ZJ,ZK,DFTLS2
  208 FORMAT(' ','REAL NS=',4F10.6)
      READ(5,500)(FMTD(I),I=1,20)
  500 FORMAT(20A4)
      WRITE(6,500)(FMTD(I),I=1,20)
      F1=N-1
      F2=J+K-2
      DO 11 I=1,NPOP
C READ POPULATION DATA FROM AN EXISTING FILE
      READ(9,FMTD)A,B
      XP(I)=A
      YP(I)=B
   11 CONTINUE
C BEGIN LOOP CALCULATING ALL STATISTICS
      DO 100 LK=1,NT
      XMUP=0.0D0
      YMUP=0.0D0
      XMP=0.0D0
      YMP=0.0D0
      SS=0.0D0
      SPX=0.0D0
      SX=0.0D0
      SY=0.0D0
      SXY=0.0D0
      SXSQ=0.0D0
      SYSQ=0.0D0
      SPY=0.0D0
      SSX=0.0D0
      SSY=0.0D0
      A12=0.0D0
      AA1=0.0D0
      AA2=0.0D0
      B1=0.0D0
      B2=0.0D0
      DO 107 I=1,NN
 800  Z=GGUBFS(DSEED)
      KK=(Z*VNPOP)+.5D0
      IF(KK.EQ.0) GO TO 800
      X(I)=XP(KK)
      Y(I)=YP(KK)
C     IF(LK.LE.5.OR.LK.GE.995)WRITE(13,108)LK,X(I),Y(I)
  108 FORMAT(' ',I5,2F7.3)
      SX=SX+X(I)
C     WRITE(6,106)SX
  106 FORMAT(' ','SX=',F14.3)
      SY=SY+Y(I)
      SXY=SXY+(X(I)*Y(I))
      SXSQ=SXSQ+(X(I)*X(I))
      SYSQ=SYSQ+(Y(I)*Y(I))
  107 CONTINUE
C     WRITE(6,120)VN
C     WRITE(6,102)SX,SY,SXY,SXSQ,SYSQ
  102 FORMAT(5F14.7)
C     WRITE(6,120)VN
  120 FORMAT(' ','VN=',F5.2)
C CALCULATE MEANS VARIANCES AND CORRELATION OF SAMPLE
      RXY=((VN*SXY)-(SX*SY))/(DSQRT(((VN*SXSQ)-(SX*SX))*
     *((VN*SYSQ)-(SY*SY))))
      ZRXY=DLOG(DSQRT((1.0D0+RXY)/(1.0D0-RXY)))
      VARX=((VN*SXSQ)-(SX*SX))/(VN*(VN-1.0D0))
      VARY=((VN*SYSQ)-(SY*SY))/(VN*(VN-1.0D0))
      SX=SX/(VN)
      SY=SY/(VN)
      DO 4 I=1,N
      XMP=XMP+X(I)
      YMP=YMP+Y(I)
C     WRITE(6,304)XMP,X(I)
    4 CONTINUE
      DO 5 I=N1,N3
      XMUP=XMUP+X(I)
C     WRITE(6,304)XMUP,X(I)
   5  CONTINUE
      DO 6 I=N2,NN
      YMUP=YMUP+Y(I)
C     WRITE(6,304)YMUP,Y(I)
   6  CONTINUE
C CALCULATE MEANS OF PAIRED AND UNPAIRED DATA
```

```
C     WRITE(6,14)N,J,K,NN,F1,F2
   14 FORMAT(4I5,2F4.1)
      XA=(XMP+XMUP)/(ZN+ZJ)
      YA=(YMP+YMUP)/(ZN+ZK)
      XMP=XMP/(ZN)
      YMP=YMP/(ZN)
      XMUP=XMUP/(ZJ)
      YMUP=YMUP/(ZK)
C     WRITE(6,304)XA,YA
  304 FORMAT(' ',2F6.3)
C CALCULATE SUMS OF SQUARES
      DO 300 I=1,N
      A12=A12+((X(I)-XMP)*(Y(I)-YMP))
      AA1=AA1+((X(I)-XMP)*(X(I)-XMP))
      AA2=AA2+((Y(I)-YMP)*(Y(I)-YMP))
  300 CONTINUE
      DO 301 I=N1,N3
      B1=B1+((X(I)-XMUP)*(X(I)-XMUP))
  301 CONTINUE
      C1=AA1+B1
      DO 302 I=N2,NN
      B2=B2+((Y(I)-YMUP)*(Y(I)-YMUP))
C     WRITE(6,824)B2,Y(I),YMUP
  824 FORMAT(' ',3F12.8)
  302 CONTINUE
C     WRITE(6,200)A12,AA1,AA2,B1,B2
  200 FORMAT(' ','SUMS OF SQUARES',1X,5F10.5)
C CALCULATE S AND S POOLED
      DO 7 I=1,N
      SS=SS+((X(I)-Y(I)-XMP+YMP)*(X(I)-Y(I)-XMP+YMP))
    7 CONTINUE
      SS=SS/F1
C     WRITE(6,15)SS
   15 FORMAT(' ','SS SQ',F9.5)
      A2=(SS*F1)/(F1-2.0D0)
      A4=((SS*SS)*(F1*F1))/(((F1-2.0D0)*(F1-2.0D0))*
     *(F1-4.0d0))
      S=DSQRT(SS)
      DO 13 I=N1,N3
      SPX=SPX+((X(I)-XMUP)*(X(I)-XMUP))
   13 CONTINUE
      DO 3 I=N2,NN
      SPY=SPY+((Y(I)-YMUP)*(Y(I)-YMUP))
    3 CONTINUE
      SP=(SPX+SPY)/(F2)
C     WRITE(6,16)SP
   16 FORMAT('0','SP SQ',F9.5)
      A1=(SP*F2)/(F2-2.0D0)
      A3=((SP*SP)*(F2*F2))/(((F2-2.0D0)*(F2-2.0D0))*
     *(F2-4.0D0))
C     WRITE(6,402)A1,A2,A3,A4
 402  FORMAT('0','THE AS',4F9.4)
      SP=DSQRT(SP)
      VVN=N
```

```
C CALCULATE PAIRED T, UNPAIRED T AND BOHJ'S TB1
      PT=(XMP-YMP-D)/(S/DSQRT(VVN))
      V1=((1.0D0/ZJ)+(1.0D0/ZK))
C     WRITE(6,17)V1
   17 FORMAT('0',F7.2)
      UT=(XMUP-YMUP-D)/(SP*DSQRT(V1))
      F=4.0D0+((A1+A2)**2.0D0/(A3+A4))
      H=(F/(F-2.0D0))/(A1+A2)
      H=DSQRT(H)
      DO 503 I=1,5
      HT(I)=H*((VL(I)*PT)+((1.0D0-VL(I))*UT))
      OHT(I)=HT(I)
      OF=F
C CALCULATE P VALUE FOR BOHJ'S TB1
      CALL MDTD(OHT(I),OF,OP(I),IER)
      IF(OP(I).LT..05)NSP5(I)=NSP5(I)+1
      IF(OP(I).LT..01)NSP1(I)=NSP1(I)+1
  503 CONTINUE
C     WRITE(6,504)(HT(I),I=1,5)
  504 FORMAT(' ',5F8.3)
C  CALC UNPAIRED T ON ALL AVAILABLE OBSERVATIONS
      DO 600 I=1,N3
      SSX=SSX+((X(I)-XA)*(X(I)-XA))
  600 CONTINUE
      DO 601 I=1,N
      SSY=SSY+((Y(I)-YA)*(Y(I)-YA))
  601 CONTINUE
      DO 602 I=N2,NN
      SSY=SSY+((Y(I)-YA)*(Y(I)-YA))
  602 CONTINUE
      SPUPTA=(SSX+SSY)/(ZN+ZJ+ZN+ZK-2)
      UPTA=(XA-YA-D)/(DSQRT((SPUPTA/(ZN+ZJ))+(SPUPTA/(ZN+
     *ZK))))
C  CALCULATE ZLS
      RR=A12/(DSQRT(AA1*AA2))
      U=(2.0D0*A12)/(AA1+AA2)
      V=A12/AA1
      W=A12/AA2
      HHAT=1.0D0/(((ZN+ZJ)*(ZN+ZK))-(ZJ*ZK*RR*RR))
      AHAT=(ZN*HHAT)*(ZN+ZK+ZJ*V)
      BHAT=(ZN*HHAT)*(ZN+ZJ+ZK*W)
C     WRITE(6,201)RR,U,V,W,HHAT,AHAT,BHAT
  201 FORMAT(' ','RR TO BHAT',1X,7F10.5)
      DSTAR=(AHAT*XMP)+((1.0D0-AHAT)*XMUP)-(BHAT*YMP)
     *-((1.0D0-BHAT)*YMUP)
  203 FORMAT(' ',F11.5)
      GHAT=(((AHAT*AHAT)/ZN)+((1.0D0-AHAT)*(1.0D0-AHAT))/ZJ)
     **(AA1/VNPT)
C     WRITE(6,203)GHAT
      GHAT=GHAT-((2.0D0*AHAT*BHAT)/ZN)*(A12/VNPT)
C     WRITE(6,203)GHAT
      GHAT=GHAT+(((BHAT*BHAT)/ZN)+(((1.0D0-BHAT)*(1.0D0
     *-BHAT))/ZK))*(AA2/VNPT)
C     WRITE(6,203)GHAT
```

```
      GHAT=DSQRT(GHAT)
      ZLS=(DSTAR-D)/GHAT
C     WRITE(6,202)DSTAR,GHAT,ZLS
  202 FORMAT(' ','ZLS STATS',1X,3F10.5)
C  CALCULATE TLS2
      AA=1.0D0/(ZN+ZJ)
      BB=1.0D0/(ZN+ZK)
      CC=(2.0D0*ZN*RR)/((ZN+ZJ)*(ZN+ZK))
      DD=DSQRT((C1+B2)/(ZN+ZJ+ZK-2.0D0))
      TLS2=(XA-YA-D)/(DSQRT(AA+BB-CC)*DD)
C     WRITE(6,204)AA,BB,CC,DD,TLS2,DFTLS2
  204 FORMAT(' ','TLS2 STATS',1X,6F10.5)
C  CALCULATE TLS4
      HA=((ZN+ZK)*AA1)/(ZN+ZJ)
      HB=((ZN+ZJ)*AA2)/(ZN+ZK)
      HC=2.0D0*A12
      H1=(ZN*(HA+HB-HC))/((ZN-1.0D0)*(ZN+ZJ)*(ZN+ZK))
      H2=(ZJ*B1)/((ZJ-1.0D0)*((ZN+ZJ)*(ZN+ZJ)))
      H3=(ZK*B2)/((ZK-1.0D0)*((ZN+ZK)*(ZN+ZK)))
      TLS4=(XA-YA-D)/DSQRT(H1+H2+H3)
      DFTLS4=((H1+H2+H3)*(H1+H2+H3))/(((H1*H1)/(ZN-1.0D0))+
     *((H2*H2)/(ZJ-1.0D0))+((H3*H3)/(ZK-1.0D0)))
C     WRITE(6,205)HA,HB,HC,H1,H2,H3,TLS4,DFTLS4
  205 FORMAT(' ','TLS4 STATS', 1X,8F10.5)
C  CALCULATE ZE
      VAR1=AA1+AA2+(1.0D0+(U*U))*(B1+B2)
      VAR2=(2.0D0*(ZN-1.0D0))+(1.0D0+(U*U))*VNUPT
      VARZE=VAR1/VAR2
      ZE1=(2.0D0*ZN*(1.0D0-U))+((ZJ+ZK)*(1.0D0-(U*U)))
      ZE2=((ZN+ZJ)*(ZN+ZK))-(ZJ*ZK*(U*U))
      ZE=(DSTAR-D)/DSQRT(VARZE*(ZE1/ZE2))
C     WRITE(6,206)VAR1,VAR2,VARZE,ZE1,ZE2,ZE,ZN
  206 FORMAT(' ','ZE STATS',1X,7F10.5)
C    CALCULATE SE
      SM=(ZN*(AA1+AA2-2.0D0*A12))/(ZN-1.0D0)
      SN=((ZJ+ZK)*(B1+B2))/(ZJ+ZK-2.0D0)
      SE=((XA-YA)*DSQRT((ZN+ZJ)*(ZN+ZK)))/DSQRT(SM+SN)
      DFSE=(((SM+SN)*(SM+SN))/(((SM*SM)/(ZN+1.0D0))+((SN*SN)
     */(ZJ+ZK)))-2.0D0)
C     WRITE(6,207)SM,SN,SE,DFSE
  207 FORMAT(' ','SE STATS',1X,4F10.5)
C CALCULATE P VALUES FOR ALL STATISTICS
      OPT=PT
      OVNPT=VNPT
      OUT=UT
      OVNUPT=VNUPT
      OUPTA=UPTA
      ODFUA=ZN+ZK+ZN+ZJ-2
      OZLS=ZLS
      OZN=ZN
      OTLS2=TLS2
      ODF2=DFTLS2
      OTLS4=TLS4
      ODF4=DFTLS4
      OZE=ZE
      OSE=SE
      CALL MDTD(OPT,OVNPT,OPPT,IER)
      ODFSE=DFSE
      CALL MDTD(OUT,OVNUPT,OPUPT,IER)
      CALL MDTD(OUPTA,ODFUA,OPUPTA,IER)
      CALL MDTD(OZLS,OZN,OPZLS,IER)
      CALL MDTD(OTLS2,ODF2,OPTLS2,IER)
      CALL MDTD(OTLS4,ODF4,OPTLS4,IER)
      CALL MDTD(OZE,OZN,OPZE,IER)
      CALL MDTD(OSE,ODFSE,OPSE,IER)
      NF=OF+.5
      NTLS4=ODF4+.5
      NDFSE=ODFSE+.5
      IF(LK.EQ.1)NLS4H=NTLS4
      IF(LK.EQ.1)NLS4L=NTLS4
      IF(NTLS4.GT.NLS4H)NLS4H=NTLS4
      IF(NTLS4.LT.NLS4L)NLS4L=NTLS4
      IF(LK.EQ.1)NSEL=NDFSE
      IF(LK.EQ.1)NSEH=NDFSE
      IF(NDFSE.GT.NSEH)NSEH=NDFSE
      IF(NDFSE.LT.NSEL)NSEL=NDFSE
      IF(LK.EQ.1)NHTL=NF
      IF(LK.EQ.1)NHTH=NF
      IF(NF.GT.NHTH)NHTH=NF
      IF(NF.LT.NHTL)NHTL=NF
      WRITE(10,506)SX,SY,VARX,VARY,XMP,YMP,XMUP,YMUP,ZRXY,
     *PT,UT,ZLS,TLS2,ZE,NF,(HT(I),I=1,5),NTLS4,TLS4,NDFSE,
     *SE,UPTA
C CALCULATE SIZE OR POWER FOR ALL STATISTICS
      IF(OPPT.LT..05)NPTPOW=NPTPOW+1
      IF(OPUPT.LT..05)NUTPOW=NUTPOW+1
      IF(OPUPTA.LT..05)NUAPOW=NUAPOW+1
      IF(OPZLS.LT..05)NPZLS5=NPZLS5+1
      IF(OPZLS.LT..01)NPZLS1=NPZLS1+1
      IF(OPTLS2.LT..05)NPLS25=NPLS25+1
      IF(OPTLS2.LT..01)NPLS21=NPLS21+1
      IF(OPTLS4.LT..05)NPLS45=NPLS45+1
      IF(OPTLS4.LT..01)NPLS41=NPLS41+1
      IF(OPZE.LT..05)NPZE5=NPZE5+1
      IF(OPZE.LT..01)NPZE1=NPZE1+1
      IF(OPSE.LT..05)NPSE5=NPSE5+1
      IF(OPSE.LT..01)NPSE1=NPSE1+1
      IF(OPPT.LT..01)NPOW2=NPOW2+1
      IF(OPUPT.LT..01)NPOW3=NPOW3+1
      IF(OPUPTA.LT..01)NPOW4=NPOW4+1
 100  CONTINUE
      WRITE(6,502)NHTL,NHTH,NLS4L,NLS4H,NSEL,NSEH
  502 FORMAT(' RANGE OF HT',I3,' TO ',I3,/,
     *'RANGE OF TLS4',I3,' TO ',I3,/,
     *'RANGE OF SE',I3,' TO ',I3)
      IF(D.EQ.0.0D0)WRITE(6,118)NPTPOW,NPOW2,NUTPOW,NPOW3,
     *NPZLS5,NPZLS1,NPLS25,NPLS21,NPLS45,NPLS41,NPZE5,
     *NPZE1,NPSE5,NPSE1,(NSP5(I),I=1,5),(NSP1(I),I=1,5),
     *NUAPOW,NPOW4
  118 FORMAT(' ',' SIZE OF PAIRED T AT .05=',I4,'  AT .01=',
     *I4,/,'SIZE OF UNPAIRED T AT .05=',I4,'  AT .01=',I4,/,
     *'SIZE OF ZLS AT .05=',I4,'  AT .01=',I4,/,
     *'SIZE OF TLS2 AT .05=',I4,'  AT .01=',I4,/,
     *'SIZE OF TLS4 AT .05=',I4,'  AT .01=',I4,/,
     *'SIZE OF ZE AT .05=',I4,'  AT .01=', I4,/,
     *'SIZE OF SE AT .05=',I4,'  AT .01=',I4,/,
     *'SIZES OF HT AT .05=',5I5,/,
     *'SIZES OF HT AT .01=',5I5,/,
     *'SIZE OF UPT-ALL AT .05=',I4,'  AT .01=',I4)
      IF(D.NE.0.0D0)WRITE(6,111)NPTPOW,NPOW2,NUTPOW,NPOW3,
     *NPZLS5,NPZLS1,NPLS25,NPLS21,NPLS45,NPLS41,NPZE5,NPZE1,
     *NPSE5,NPSE1,(NSP5(I),I=1,5),(NSP1(I),I=1,5),
     *NUAPOW,NPOW4
  111 FORMAT(' ',' POWER OF PAIRED T AT .05=',I4,' AT .01=',
     *I4,/,'POWER OF UNPAIRED T AT .05=',I4,' AT .01=',I4,/,
     *'POWER OF ZLS AT .05=',I4,'  AT .01=',I4,/,
     *'POWER OF TLS2 AT .05=',I4,'  AT .01=',I4,/,
     *'POWER OF TLS4 AT .05=',I4,'  AT .01=',I4,/,
     *'POWER OF ZE AT .05=',I4,'  AT .01=', I4,/,
     *'POWER OF SE AT .05=',I4,'  AT .01=',I4,/,
     *'POWER OF HT AT .05=',5I5,/,
     *'POWER OF HT AT .01=',5I5,/,
     *'POWER OF UPT-ALL AT .05=',I4,'  AT .01=',I4)
  506 FORMAT(14F16.12,I3,5F16.12,2(I3,F16.12),F16.12)
      WRITE(6,703)DSEED
  703 FORMAT(' ',' FINAL SEED=',D16.9)
      WRITE(12,850)DSEED
      STOP
      END
```